OPEN ACCESS

University of BRISTOL

This electronic thesis or dissertation has been downloaded from Explore Bristol Research, http://research-information.bristol.ac.uk

Author:
Spooner, Alice L. R

Title:
Investigating and assessing comprehension ability

General rights
Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy
Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

• Your contact details
• Bibliographic details for the item, including a URL
• An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.

Investigating and assessing comprehension ability

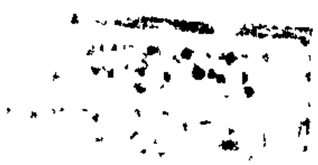
Alice L. R. Spooner

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Science.

Department of Experimental Psychology

December 2001

Word count: 38, 736



ABSTRACT

In the first section of the work presented here, a sentence recognition test (c.f., Oahkill 1982, Oakhill, Yuill & Parkin, 1986) was administered under high and low memory demands, to seven- and eight-year-old skilled and less-skilled comprehenders. The groups did not differ in performance on the test, providing no evidence that poor integration of information or poor memory for story information lie at the root of comprehension deficits. Further findings suggested, however, that the sentence recognition test may have been insensitive to cognitive differences between the groups.

An additional explanation for the null results was that the Neale Analysis of Reading Ability (Neale, 1958; Neale, 1997; Neale, Christophers, & Whetton, 1989), on which children's comprehension ability had been assessed, had overestimated the comprehension deficit of the less-skilled groups. The second section of the work presented in this thesis, therefore, investigated how performance on various comprehension measures was affected by test demands. It was found that open-ended questions – such as those used in the Neale Analysis of Reading Ability – proved problematic for some children, who did not otherwise exhibit comprehension difficulties. It was also demonstrated that children with poor reading accuracy obtained lower comprehension scores on the Neale Analysis of Reading Ability than was predicted from their performance on a listening comprehension test. In addition, it was found that while word recognition and listening comprehension were independent predictors of unaided reading comprehension, the measures of reading accuracy and reading comprehension provided by the Neale Analysis of Reading Ability were not independent. While it was acknowledged that the Neale Analysis of Reading Ability is a useful tool in education, it was suggested that this test should not be used an analytic tool of assessment.

ACKNOWLEDGEMENTS AND DEDICATION

I acknowledge the honour that it has been to have worked with Sue Gathercole and Alan Baddeley, and I thank them for their high-calibre training and continued support. Thanks also to Chris Jarrold, for patient advice and explanations on a number of occasions. This work could not have been completed without the assistance of the Bristol schools, in particular Holy Cross, South Street, and Our Lady of the Rosary, who agreed to my multiple returns with friendliness and interest, and whose children were a delight to work with. I thank my parents – Terry and Elfrida Spooner – for giving me the brain matter and stubborn dedication necessary for producing a thesis, and for – in my mother's wise words – never pushing me, but running alongside me cheering. I also praise my sisters – Merrill and Georgia – for their continual entertainment of everyone and their magnificent adoration of each other.

This thesis is dedicated to Katy Lobley

for her outstanding and most treasured comprehension.

AUTHOR'S DECLARATION

I declare that the work contained within this thesis was carried out in accordance with the Regulations of the University of Bristol. The work is original and no part of the dissertation has been submitted for any other degree. Any views expressed in the dissertation are those of the author and in no way represent those of the University of Bristol. The thesis has not been presented to any other University for examination either in the United Kingdom or overseas.

Signed....A Spooner.....Date..08 July 2002

CONTENTS

	Page
CHAPTER 1	
INTRODUCTION	**1**
1.1 Comprehension	**1**
1.1.1 Mental models	2
1.1.2 Construction-integration	3
1.2 Processes in comprehension	**5**
1.2.1 Integration of information	6
1.2.2 Inference generation	7
1.2.2.1 Minimalist hypothesis	8
1.2.2.2 Constructionist theory	9
1.2.3 Memory	10
1.3 Poor comprehension	**15**
1.3.1 Semantic deficits in poor comprehension	16
1.4 Poor comprehension and processes in comprehension	**17**
1.4.1 Integration of information	18
1.4.2 Inference generation	20
1.4.3 Memory	23
1.5 Summary of introduction and overview of thesis	**25**
CHAPTER 2	
POOR COMPREHENSION, INTEGRATION AND MEMORY	**27**
2.1 Experiment 1	**29**
2.1.1 Method	29
2.1.2 Results	32
2.1.3 Discussion	33
2.2 Experiment 2	**33**
2.2.1 Method	33
2.2.2 Results	35
2.2.3 Discussion	37

2.3 Experiment 3 38
2.3.1 Method 38
2.3.2 Results 39
2.3.3 Discussion 41
2.4 Experiment 4 42
2.4.1 Method 42
2.4.2 Results 43
2.4.3 Discussion 46
2.5 Experiment 5 48
2.5.1 Method 48
2.5.2 Results 50
2.5.3 Discussion 53
2.6 General Discussion 54

CHAPTER 3
METHODOLOGICAL MANIPULATIONS OF COMPREHENSION 58
3.1 Experiment 6 62
3.1.1 Method 62
3.1.2 Results 65
3.2 Experiment 7 72
3.2.1 Method 72
3.2.2 Results 73
3.3 General Discussion 76

CHAPTER 4
READING AND COMPREHENSION AS SEPERABLE COMPONENTS OF THE NEALE ANALYSIS OF READING ABILITY 82
4.1 Development of the listening comprehension and unaided reading comprehension tests 86
4.1.1 Design of materials 86
4.1.2 Development of materials 88

4.2 Experiment 8 93
4.2.1 Method 94
4.2.2 Results 94
4.2.3 Discussion 97
4.3 Experiment 9 98
4.3.1 Method 98
4.3.2 Results 99
4.3.3 Discussion 101
4.4 Experiment 10 104
4.4.1 Method 105
4.4.2 Results 105
4.4.3 Discussion 114
4.5 General Discussion 117

CHAPTER 5
GENERAL DISCUSSION 121
5.1 Poor comprehension, integration and memory 121
5.1.1 The sentence recognition test 121
5.1.2 The Neale Analysis of Reading Ability comprehension assessment 122
5.1.3 Poor comprehension, integration and memory – summary 125
5.2 The Neale Analysis of Reading Ability reading assessment 125
5.3 Comprehension ability as opposed to test performance 127
5.3.1 Reading condition 127
5.3.2 Test versus independent class work 127
5.3.3 The Neale Analysis of Reading Ability versus alternative tests 129
5.3.4 Comprehension ability versus performance – summary 129
5.4 Future comprehension test 130
5.5 Future directions 130

REFERENCES 133

APPENDIX A 138

APPENDIX B 142

TABLES

page

Table 2.1 30

Participant characteristics in Experiment 1. Ages in years, standard deviations in parentheses.

Table 2.2 32

Mean number of correct recognitions and rejections in Experiment 1. Standard deviations in parentheses.

Table 2.3 35

Mean number of correct recognitions and rejections in Experiment 2. Standard deviations in parentheses.

Table 2.4 39

Mean number of correct recognitions and rejections in Experiment 3. Standard deviations in parentheses.

Table 2.5 43

Participant characteristics for Experiment 4, by order of presentation condition. Ages in years, standard deviations in parentheses.

Table 2.6 44

Mean number of correct recognitions and rejections in Experiment 3, by order of presentation condition. Standard deviations in parentheses.

Table 2.7 49

Participant characteristics in Experiment 5, by order of presentation condition. Ages in years, standard deviations in parentheses.

Table 2.8 50

Mean number of correct recognitions and rejections in Experiment 5, by order of presentation condition. Standard deviations in parentheses.

Table 3.1 63

Participant profiles in Experiment 6. Ages in years. Standard deviations in parentheses.

Table 3.2 66

Mean comprehension scores in Experiment 6, with standard deviations in parentheses.

Table 3.3 67

Summary of performance by children who completed Levels 1 to 4 on *aloud tell me* and *aloud true/false* in Experiment 6. Percentage correct, with standard deviations in parentheses.

Table 3.4 74

Mean level attained by each group, and mean comprehension scores attained in each reading condition in Experiment 7. Standard deviations in parentheses.

Table 4.1 89

Scores from pilot stage 1 of the comprehension test development, and number of children attempting each story.

Table 4.2 90

Scores from pilot stage 2 of comprehension test development. Maximum score at Level 1 was 4, maximum score at Levels 2 to 4 was 8.

Table 4.3 93

Mean number of questions correct in pilot stages 3 and 4 of comprehension test development, standard deviations in parentheses.

Table 4.4 97

Contingency table of scores in Experiment 8.

Table 4.5 100

Contingency table of scores in Experiment 9.

Table 4.6 106

Mean scores attained by the sample of children in Experiment 10. Ages in years. Standard deviations in parentheses.

Table 4.7 108

Mean performance of groups as selected by the Neale Analysis of Reading Ability in Experiment 10. Ages in years. Standard deviations in parentheses.

Table 4.8 109

Number of children in each reading group achieving three profiles across the Word Decision Test and the listening comprehension test, in Experiment 10. Also expressed as percentage of each group.

Table 4.9 113

Hierarchical multiple regression analyses conducted in Experiment 10, with unaided reading comprehension predicted as the dependent variable.

Table 4.10 114

Hierarchical multiple regression analyses conducted in Experiment 10, with unaided reading comprehension predicted as the dependent variable.

CHAPTER 1

INTRODUCTION

1.1 Comprehension

Many definitions of comprehension appear both elusive and intuitive, hardly bringing us closer to what words such as comprehension, understanding and meaning really mean. Johnson-Laird described understanding as 'having a "working model" of the phenomenon in your mind' (Johnson-Laird, 1985, p2). This introduces the explicit awareness that understanding (and its sister, comprehension) happen in the mind, and that they involve some kind of representation of information.

Throughout this thesis, comprehension shall be defined as the 'formation of a correct conceptual representation of presented information'. Three words are crucial to this definition – *correct*, *conceptual* and *presented*. Firstly, the word *presented* makes the difference between comprehension and understanding. When we comprehend text or speech, or indeed any other form of communication, we are being presented with a message, and comprehension is the end-point of a relay in communication. We are not concerned here with understanding, which can occur independent of external input, as the formation of any idea. Both understanding and comprehension refer to the possession of conceptual representations of information, but comprehension is restricted to the representation of information that has been presented for the purpose of communication.

Secondly, the word *conceptual* necessitates some kind of interpretation to be conducted upon the information. Consider the example of simple repetition, which certainly involves representation of the information, in auditory form. Repetition, therefore, would satisfy the definition of comprehension as the 'formation of a correct representation of presented information'. The difference between the representation of surface information and comprehension is parallel to that between perception and conception.

Lastly, the word *correct* in the above definition makes the distinction between comprehension and miscomprehension. If the representation is not correct, than neither is the comprehension. This thesis is concerned with the accuracy of comprehension – any old representation does not count. One complication is the fact that many levels of comprehension may be derived from one situation. Consider the example of children being administered a comprehension test. They might construct comprehension of individual phrases, or of the text as a whole, or go beyond that to interpret what the experimenter's mood is, or the intention underlying the experimental design. However, that complication was avoided in the work presented in this thesis, by specifying to children that they would be asked questions about what happened in stories. Successful comprehenders in this work were therefore those children who formed correct representations of what happened in stories.

In this chapter, a framework shall be described, which can account for features of comprehension. Some of the processes shown to underlie comprehension shall then be discussed, followed by evidence that individuals with low comprehension performance also exhibit low performance on tasks that tap these underlying processes.

1.1.1 Mental models

Incoming written text or discourse (passages of speech) are not stored as the surface data (the sequence of words), nor as individual propositions (the low level concepts). Rather, the information is stored as a model of the situation that the words were assembled to describe. These models are referred to by Johnson-Laird (Johnson-Laird, 1985) as mental models, and by others as situation models, and they correspond to what was referred to in the definition of comprehension provided earlier as 'conceptual representations'.

Johnson-Laird described mental models as being 'constructed from tokens arranged in a particular structure to represent a state of affairs' (Johnson-Laird, 1985, p398). The structure of the mental model represents the structure of the situation described. If, for example, a spatial array is described, the information

is represented in the mind as a spatial array, not a set of grammatical propositions. Furthermore, not only would this spatial array be represented in a spatial manner, but the set of relative locations of individual objects relative to other individual objects would be assembled to form a global model of all the objects in the array described. For example, consider the two following propositions, *A is above B* and *B is above C*, and accept that we know that B refers to the same item in each proposition. A mental model is constructed of the relationships between these three items, with B both below A and above C. We, therefore, cannot help but know that A is above C, because it is so in the mental model.

A wealth of evidence has demonstrated that mental models are indeed structured in this manner. Sentence recognition tests have demonstrated that people do not retain information about the exact words or word order, but rather retain the meaning – the mental model (e.g., Bransford, Barclay, & Franks, 1972; Bransford & Franks, 1971; Sachs, 1967). For example, given the original sentence *He sent a letter about it to Galileo, the famous Italian scientist*, people would falsely recognise the sentence *He sent Galileo, the great Italian scientist, a letter about it* (Sachs, 1967). Furthermore, sentence recall tests have shown that the meaning that is retained contains more than was explicitly provided by the text. For example, having learned the sentence *Nurses are often beautiful*, recall of the sentence was improved by the cue *actress* (Anderson & Ortony, 1975) – the mental model incorporated other knowledge in its interpretation of text information. In addition, mental models are updated with incoming information during comprehension, (Morrow, Bower, & Greenspan, 1989).

1.1.2 Construction-integration

Mental model theory provides an account of what kind of representation is formed in comprehension. The construction-integration model (Kintsch, 1988, but also see Kintsch & van Dijk, 1978), provides an account of how the representations are formed.

The model assumes that word meanings and other knowledge exist in an associative network. This means that when any item is processed, associated items are also activated. The term associated means that if items X and Y are frequently experienced together, and therefore activated together, a connection will form between the representations of X and Y, so that activation of one will activate the other. The name given to the construction-integration model highlights the two processes that are fundamental to the model. Firstly, text-base representations are constructed through spreading activation from existing knowledge and the linguistics of the text or discourse – words, syntax, grammar. Representations are constructed in this manner in chunks that correspond to the amount of text information that can be held in working memory. The separate textbase activations are then integrated though overlapping activations to form a coherent representation of the text content, which also incorporates activations that were not mentioned in the text but are associated though world knowledge. The resulting coherent representation can be viewed as a mental model of the text.

The construction-integration model can explain how mental models are formed. It can also account for the experimental evidence. Consider the Galileo example (Sachs, 1967) described in section 1.1.1. People claimed that they had heard sentences before that they had not, in fact, heard before, but which shared meaning with original sentences. People had stored the meanings of sentences but not the actual sentences. In the construction-integration model, the final representations are abstracted from the text information, and stored as an activation in meaning, not in words, syntax or grammar. The surface information of text or discourse is discarded once meaning has been accessed.

The construction-integration model can also account for the storage of information not explicitly mentioned in text. Comprehension exists as a pattern of activation across the associative network, and associated knowledge is also activated and thereby joins the pattern. For example, in the work by Anderson and Ortony (1975) described in Section 1.1.1, the cue *actress* aided recall of the sentence *Nurses are often beautiful*. The connection between the cue and the

sentence lay in the word *beautiful*. When *beautiful* was activated in the associative network, its close associate *actress* was also activated. *Actress*, therefore, was connected to the pattern of activation that corresponded to the sentence *Nurses are often beautiful*. When then given the cue, *actress* was activated in the network, also activating the representation of *Nurses are often beautiful*. In addition, the model can account for the flexibility of mental models. Section 1.1.1 mentioned how mental models adapt to incoming information, remaining consistent with the current state of affairs described by the text or discourse (Morrow et al., 1989). The construction-integration model describes a flexible system, in which patterns of activation are continually altered by the activation of individual items and their associates.

To summarise, comprehension is defined for the purposes of this thesis as the successful formation of a correct conceptual representation of presented information. Within comprehension, the theory of mental models provided by Johnson-Laird (1985) provides a description of the conceptual representations formed in comprehension. The construction-interaction model proposed by Kintsch (Kintsch, 1988; Kintsch & van Dijk, 1978) provides an account of how these representations are formed.

1.2 Processes in comprehension

The work presented here is concerned with the processes that underlie comprehension, between abstraction of propositions to construction of global mental representations. In this section, two of the most important processes – integration of information and inference generation – and the capacity that is necessary for conducting these processes – memory – are discussed relative to the framework presented in Section 1.1.

Throughout this thesis, the word *text* will inevitably be used to refer to both written and spoken material. Where written text or spoken passages needs to be specified, they will be referred to as *written text* or *discourse* respectively. Similarly, the term *comprehension* shall be used to refer to general comprehension, regardless of modality. Where it is necessary to specify

modality, the terms *reading comprehension* or *listening comprehension* shall be used. For the most part, the work presented here explored general comprehension, and the processes that this work was concerned with take place after written words and speech sounds have been transformed into textbase representations.

1.2.1 Integration of information

The construction-integration model of comprehension states that information at different levels are integrated to form a coherent global representation of the text information. Without sufficient integration, the text is represented as abstract and isolated items of information. Integration is the crucial last step.

Bransford and Franks (1971) investigated the extent to which information is integrated between sentences. Undergraduates listened to a number of sentences, which contained one, two or three propositions that could be integrated into a coherent whole of four propositions. For example, the three sentences *The ants were in the kitchen*, *The ants ate the sweet jelly*, and *The ants in the kitchen ate the jelly which was on the table* can be integrated to form the sentence *The ants in the kitchen ate the sweet jelly which was on the table.* Participants falsely recognised sentences that they had not heard before but which could be created by integrating propositions that they had heard before. This finding demonstrated that people do indeed integrate information across propositions.

Similar work was conducted by Paris and Carter (1973) with children, aged seven- and ten-years. Children listened to sentences, and were told to remember the sentences. The sentences came in sets of three, which could be integrated to form a global situation, for example, *The bird is inside the cage*, *The cage is under the table*, *The bird is yellow*. Children were later tested for their recognition of these sentences. They were very likely to falsely recognise sentences that they had not heard before but which concurred with the global situation, such as *The bird is under the table*. This finding demonstrated that children, too, integrate information across propositions.

1.2.2 Inference generation

Inferences are items of information that are not stated in text, but are generated in response to textual information. Some of these inferences are automatic processes required to link referents and propositions, in order to construct meaning at the shallow level of deconstructing grammar. Other inferences link low-level meanings into higher-level meanings, bringing the whole text under one concise summary. In the construction-integration model, these inferences are made through associative activation in the knowledge net, and the concise summary corresponds to the mental model. Inferences are not all restricted to events within the text, but might consider what the author felt when writing it, or what the text's purpose could be. Research tends to focus on narrative text because it is argued (e.g., Graesser, Singer, & Trabasso, 1994) that they tap our knowledge of the world in a manner that more informational texts do not; the situations described in such texts are more alike to everyday situations, of which we are familiar in constructing mental models. Of the inferences that are generated around explicit text information, some are generated on-line during text comprehension in order to make sense of the text, and others are made off-line, after reading, during a subsequent task such as remembering or perhaps answering a freshly-considered query about the text.

It is not agreed how frequently automatic on-line inferences are generated under normal reading conditions. For example, the idea of explicit textbase processing holds that no inferences are made, so that the reader represents nothing beyond what is stated explicitly within text. At the other extreme, promiscuous inference generation proposes that all types of inference are generated on-line, so that the reader constructs a full life-like mental representation of the situation. However, there are two main positions on what types of inferences are made on-line in the construction of the mental model during initial comprehension, and both positions can satisfy the construction-integration model. The minimalist hypothesis proposes that the only on-line inferences to be generated are those that are easily accessible from explicit statements or world-knowledge, and those that are necessary to make sense of explicit statements. Constructivist theory, on the other hand, proposes that additional

inferences are generated in order to achieve global coherence and to explain why certain information is mentioned.

1.2.2.1 Minimalist hypothesis

The minimalist hypothesis proposed by McKoon and Ratcliff (1992) argues that the only inferences to be generated automatically during reading comprehension are those that are based on easily available information, be it explicit statements or world-knowledge, or those that are necessary for local coherence. The term local coherence used by McKoon and Ratcliff (1992) refers to the attempt to make sense of explicit statements that exist in working memory at the same time, and the term local inferences are those inferences necessary for local coherence. In effect, this means linking propositions that are one or two sentences apart. By global inferences, on the other hand, McKoon and Ratcliff (1992) refer to the inferences made across text, in order to link together isolated meanings, and they claim that global inferences are only made when sections of text are not in themselves locally coherent.

McKoon and Ratcliff (1992) provide evidence that they claim demonstrates automatic encoding of local, but not global, inferences. It is not, however, clear how this evidence concurs with evidence mentioned earlier. For example, in the work by Anderson and Ortony (1975) described in Sections 1.1.1 and 1.1.2, the cue actress aided recall of the sentence *Nurses are often beautiful*, demonstrating that people do encode more than appears necessary to make sense of information explicitly stated in text.

McKoon and Ratcliff (1992) argue that local, but not global, inferences are made automatically in comprehension. If their interpretation of the term global corresponds with the level of mental models, then their minimalist view is inconsistent with that presented so far. However, in order to test global representations of story information McKoon and Ratcliff (1992) tested availability of story goals. It could be argued that story goals are a higher, more abstracted level than mental models which represent the event situation. It

seems, therefore, that the minimalist hypothesis might yet agree with the construction-integration model and the theory of mental models.

1.2.2.2 Constructionist theory

The constructionist theory proposed by Graesser et al. (1994) embraces the principle of search (or effort) after meaning, described by Bartlett (1932). This principle makes three assumptions about the manner in which the reader attempts to make sense of text: The reader goal assumption asserts that the reader attempts to construct a meaning representation that addresses the reader's own goal. The coherence assumption predicts that the reader attempts to construct a meaning representation that is both locally and globally coherent. Local coherence refers to the meaningful organisation of elements within small sections of text, whilst global coherence is the process by which local chunks are integrated to form higher order meaning. The third assumption is the explanation assumption, by which the reader attempts to explain why items of information are mentioned. Together, these three assumptions lead to a model of comprehension in which the reader works to construct a meaning representation of text in which information is given for a reason, sense is made in subsections of text, and that these subsections fit together to form a more overall meaning, which suits the readers own needs.

Constructionist theory predicts that during normal comprehension conditions, six classes of inferences are generated on-line: referential inferences, case structure role assignment and causal antecedent inferences are required for local coherence, and inferences regarding superordinate goals, themes and character emotion reactions are necessary for global coherence. In addition, causal antecedent and superordiante goal inferences enable event explanation. Together, these six types of inferences enable successful search after meaning. Graesser et al. (1994) propose that the minimalist hypothesis predicts that only the first three classes of inferences should be generated, and present evidence that all classes except thematic inferences are usually generated on-line. Of particular interest is their finding that inferences about goals are generated, in direct contrast to evidence described by McKoon and Ratcliff (1992), which

found that inferences about goals are only generated when it is necessary for local coherence.

In summary, two theories of inference generation during comprehension have been described. They disagree on the extent to which inferences are generated. In particular, the minimalist hypothesis states that only local inferences are normally encoded, whilst the constructionist theory states that both local and global inferences are generated. Evidence does appear to suggest that more is encoded than just that explicitly stated in text and necessary for coherence at the local level.

1.2.3 Memory

The construction-integration model of comprehension (Kintsch, 1988) described in Section 1.1.2 proposes that text information is processed in cycles. During each cycle, new information is held in working memory as a chunk of propositions. Each cycle ends with the new chunk of information being assimilated with existing information, either directly through coherence, or by the use of inferences and integration. Individuals with greater working memory capacity, it follows, would be able to store larger chunks of propositions, in fewer cycles, so process the incoming information more efficiently than individuals with lower working memory capacity. It follows that comprehension ability might vary with differences in working memory. Research into relationships between working memory and comprehension has focussed on the storage of phonological information on the phonological loop of the working memory model proposed by Baddeley (Baddeley, 1986; Baddeley, 1996) and on the concurrent storage and processing of information. The research in these two areas shall be discussed in turn.

In reading or listening comprehension, words are decoded and heard in strict sequential order. But grammar does not work sequentially. Some sentence structures cannot be resolved until the end of the sentence, and sometimes people make the wrong interpretation and have to backtrack to correct themselves. In these situations it might appear particularly useful for recent

surface text information to have been stored for referral on the phonological loop.

One method of assessing the contribution of the phonological loop to a task is to examine the effect of articulatory suppression on performance on that task. Articulatory suppression refers to the production of irrelevant speech (such as *the the the*) during a task. Such verbal production is supposed to implement the phonological loop, and it follows that if articulatory suppression impairs performance on a task, the task also taps the phonological loop (Estes, 1973; Levy, 1971; Murray, 1967). In work by Baddeley, Eldridge and Lewis (1981), a sentence verification test was used to examine the use of the phonological loop during sentence comprehension. Adults were required to identify whether sentences were semantically anomalous, with or without articulatory suppression. Articulatory suppression impaired ability to identify sentences that were semantically anomalous, but did not affect identification of sentences that were not semantically anomalous. This suggests that the phonological loop was implemented in the processing of sentences that required some kind of checking, but not otherwise involved. This might be accounted for by comprehension being a mostly automated process in adults, with the need for problematic sections to be processed through reference to the phonological store of text.

A method of assessing phonological loop capacity is to administer simple span tests. These require participants to repeat back verbal items (e.g., words or digits) in the order that they were presented. Since these items are supposed to be maintained on the phonological loop, the number of items that an individual can correctly recall provides a measure of their phonological working memory capacity. Dixon, LeFevre and Twilley (1988) examined the correspondence between performances on a multiple-choice comprehension test (the Nelson-Denny), a comprehension test that examined ability to make inferences in text, and word and digit recall. Dixon et al. (1988) found no correspondence between performances on the comprehension tests and the simple span tests, suggesting that phonological working memory capacity does not influence comprehension

performance. This in turn indicates that the phonological loop is not normally involved in comprehension.

Studies with children, on the other hand, have suggested that the phonological loop is indeed involved in their comprehension. Gathercole, Willis, Baddeley and Emslie (1994) examined the relationship between performance of four- and five-year-olds on the Test for Reception of Grammar (TROG) (Bishop, 1989), the Children's Test of Non-Word Repetition (the CNRep) (Gathercole et al., 1994), and digit recall. The TROG assesses children's ability to correctly identify, from a selection of four, the picture that correctly portrays a situation described in a sentence, and therefore examines sentence comprehension. The CNRep assesses children's ability to correctly recall non-words, and is therefore a simple span task, as is the digit recall task. Gathercole et al. (1994) found a small but significant correlation between performance on the TROG and both the CNRep and digit recall task. Similarly, Adams, Bourke and Willis (1999) found a correlation between four- and five-year-olds' performance on a listening comprehension test and on both word and digit recall tests. These findings indicate that young children's comprehension ability is related to their phonological working memory capacity, suggesting that the phonological loop is indeed implemented in children's comprehension.

The second area of research into the relationship between memory and comprehension has explored the ability to store information whilst also processing information. This ability is assessed by complex span tests, as first devised by Daneman and Carpenter (1980). An example of such a test is the listening span test. Participants listen to a series of sentences, and conduct some processing on each sentence such as a sensibility judgement. The participant is then required to recall the last word of each sentence. Daneman and Carpenter (1980) proposed that performance on complex span tests would provide a better predictor of comprehension ability than performance on simple span tests because comprehension also requires concurrent processing and storage of information, rather than storage alone. Indeed, a number of correlational studies have found this pattern. Daneman and Carpenter (1980) found that performance

on a complex span task (in this case, reading a series of sentences aloud with last word recall) correlated more highly with a number of different measures of comprehension ability than performance on a simple span task (word list recall). Similarly, Dixon et al. (1988) found that comprehension correlated with reading span (sensibility judgement plus last word recall) but not with digit span (simple list recall). Furthermore, a meta-analysis conducted by Daneman and Merickle (1996) on 77 studies revealed that performance on tasks that require concurrent processing and storage provide better predictors of comprehension ability than tasks that require storage alone.

The relationship between complex working memory and comprehension ability also appears to extend to children. Seigneuric, Ehrich, Oakhill and Yuill (2000) investigated the correlation between comprehension and performance on five complex span tasks, in eight-, nine- and ten-year-olds. One task required children to complete the final words of sentences and recall the final words. Another required children to spot the odd-one-out in sets of words, and recall the odd words. In a digit task children read aloud series of digits, to recall the final digit of each set, and in another to spot the largest number from sets and recall the largest numbers. In the spatial tasks, children completed a line on a naughts-and-crosses grid, and recalled the line positions. Comprehension ability correlated highly with all complex span tasks except the spatial task.

Complex working memory appears to be a better predictor of comprehension ability than simple working memory in children as well as in adults, but only past a certain age. In their work with seven-year-olds, Leather and Henry (1994) administered a reading comprehension test, a simple span task and a listening span test (complex span), in which children filled in the final words of a series of sentences and recalled those final words in order. Whilst they found a significant correspondence between both simple span and complex span and comprehension ability, the relationship was stronger with the complex span task. Working with younger children, Adams et al. (1999) administered a battery of working memory tasks to four- and five-year-olds, including word and digit list recall (simple span) and the same listening span test administered

by Leather and Henry (1997). Adams et al. (1999) found equivalent correlations between performance on each of these tasks and comprehension. Whilst this may at first seem not to agree with the findings so far presented, simple span tasks may make processing demands on young children that are automated in slightly older children and adults. Therefore, what are referred to as simple span tasks for adults because they make no explicit processing demand might in fact make complex working memory demands on young children.

In summary, evidence suggests that simple storage of phonological information is not typically a factor in the normal variation in adult comprehension ability. It does, though, appear that storage capacity relates to comprehension ability in difficult circumstances, suggesting that the phonological loop might provide a back up of text information which is referred to when usual comprehension procedures fail. Instead, adult comprehension ability appears to relate to the ability to store and process information at the same time. Child comprehension appears to relate to both storage ability and concurrent storage and processing ability. One interpretation of this finding might be that the phonological loop is used in comprehension by children whilst not by adults, because the processes used in adult comprehension are not yet fully efficient, and so children rely more on off-line processing. Another account might be that processes required by simple span tasks might make demands on children that are equivalent to the demands made on adults by the complex span tasks.

One last point is that the demonstration of some correlational relationship between comprehension ability and memory does not demonstrate the direction of the relationship. That comprehension relates to ability to store and process information might not be that surprising, since comprehension itself requires concurrent storage and processing of information. The demonstration might, therefore, be a circular one, or, alternatively, the demonstration of common factors underlying both comprehension and memory.

1.3 Poor comprehension

As with all cognitive abilities, comprehension varies across the population. The research reported in this thesis was concerned with those individuals whose comprehension abilities were below what would have been expected from other aspects of their cognitive profiles, and investigated possible explanations for this deficit. The work concentrated solely on children in their third year at school, aged seven and eight. These children are referred to throughout as less-skilled comprehenders.

Much of the work on less-skilled comprehension has approached comprehension from the area of reading, and this thesis begins by adopting the view of comprehension as a component of reading ability. Most research in the field of poor comprehension, therefore, considers some measure of the individual's reading skill, as measured by ability to access the meanings or sounds of written words, and this thesis also follows that tradition.

Work has indicated that text and discourse comprehension are components of a general comprehension ability (Carr, Brown, Vavrus, & Evans, 1990; Nation & Snowling, 1997; Nation & Snowling, 1998b; Stothard & Hulme, 1992) so that children with poor text comprehension also exhibit poor discourse comprehension. This thesis therefore begins by assuming generality of comprehension across modality, and infers that children selected as skilled or less-skilled comprehenders on the basis of their reading comprehension provide a reliable representation of children with general good and poor comprehension ability. Chapter 4 later examines this assumption.

Section 1.3.1 reviews the work conducted by Nation, which provides a profile of the semantic deficits exhibited in poor comprehension, illustrating what poor comprehension might mean in terms of actual linguistic ability. Section 1.4 then reviews evidence that poor comprehenders have some deficit in each of the component processes introduced in Section 1.2.

1.3.1 Semantic deficits in poor comprehension

The work reviewed in this section matched groups of skilled and less-skilled comprehenders on chronological age, single non-word reading ability, and often non-verbal ability. The groups differ on either reading or listening comprehension, as measured by ability to answer questions about stories that they have either read aloud themselves or listened to.

Nation and Snowling (2000) presented skilled and less-skilled reading comprehenders with the words from sentences such as *The donkey kicked the door* in random word orders. The less-skilled comprehenders were less able than the skilled comprehenders to rearrange the words into the original meaningful sentences. This may reflect the superior comprehension of the skilled comprehenders, enabling them to construct meaning from words even without correct syntactical information. It may alternatively reflect reduced awareness of correct word order in less-skilled comprehenders. In either case, the results demonstrate that skilled and less-skilled comprehenders exhibit differences at the syntactic level, even with simple sentences.

Evidence also suggests that comprehension groups differ in their semantic skills, even at the level of single word meaning. Nation, Marshall, and Snowling (2001) found that less-skilled comprehenders were slower and less accurate than skilled comprehenders at a simple picture naming task. The less-skilled comprehenders were especially impaired at producing low-frequency words. In addition, Nation and Snowling (1998a) found that less-skilled comprehenders were also slower and less accurate than skilled comprehenders at making synonym (e.g., big – huge) judgements about heard word-pairs, exhibiting particular difficulty with low-imageability items. This finding suggests that the semantic difficulties that poor comprehenders exhibited in picture naming were not restricted to production. Furthermore, whilst less-skilled comprehenders produced fewer semantic associations to cue words (e.g., big – huge, large) than skilled comprehenders, they did not produce fewer rhyming associations (e.g., mat – cat, hat), demonstrating that their difficulties with word tasks did not extend to phonological manipulations. Commensurate

with this latter pattern of results, Nation, Adams, Bowyer-Crane, and Snowling (1999) found that, whilst skilled and less-skilled comprehenders exhibited comparable ability to recall lists of concrete words, such as *tooth*, less-skilled comprehenders were less able to recall lists of abstract words, such as *pride*. Less-skilled comprehenders appear to have reduced access to the meanings of single words.

Finally, less-skilled comprehenders appear less able than skilled comprehenders to apply context to identifying words in text. Nation and Snowling (1997) selected groups of skilled and less-skilled comprehenders on the basis of their performance on a listening comprehension test. Although the groups were matched for non-word reading ability, they differed in their text reading accuracy. This suggests that the skilled readers were able to use the surrounding context to identify difficult words, whilst less-skilled comprehenders were less able to do so. Also selecting groups on the basis of listening comprehension, Nation and Snowling (1998b) matched comprehension groups on single word reading ability, and examined the speed and accuracy of word identification following context in relation to reading the word in isolation. Skilled comprehenders exhibited more contextual facilitation than less-skilled comprehenders, in both speed and accuracy. These findings demonstrate that the comprehension limitations of less-skilled comprehenders can adversely affect not only their grasp of the subject material, but also their ability to resolve difficulties in reading.

1.4 Poor comprehension and processes in comprehension

The work reviewed in Section 1.2 employed two strategies for investigating which skills underlie successful comprehension. The first strategy was to design materials that reveal, in the types of errors most commonly made, the processes that had been utilised in comprehending those materials. This type of research has revealed that information is integrated and that inferences are generated during comprehension. The second strategy investigated how the natural spread of comprehension skill across populations related to the spread of memory ability. The findings from this area of research suggest that individuals with

better comprehension ability also tend to have better memory ability, inferring that memory is indeed a component process of comprehension.

Section 1.4 reviews work that has compared the general proficiency in integration, inference and memory skills between groups of individuals who differ in comprehension ability. By comparing children with different comprehension abilities on tasks known to tap certain skills, the contribution of those skills to comprehension can be investigated. Section 1.4 reviews evidence that impairments in the three processes described in Section 1.2 might underlie poor comprehension.

1.4.1 Integration of information

Section 1.1.2 described the construction-integration framework of comprehension, and Section 1.2.1 reviewed the evidence that indicates that integration of information is indeed an important process in comprehension. It is an additional topic whether some deficiency in integrating information might lie at the root of poor comprehension.

In two studies, Oakhill (1982) and Oakhill, Yuill, and Parkin (1986) investigated the tendency of seven- and eight-year-old skilled and less-skilled comprehenders to integrate information. When asked to identify which sentences they had heard in a short story, children found it difficult to reject sentences that they had not heard before but which shared a meaning with story information – semantically congruent foils. This did not simply reflect low accuracy or low confidence in rejecting sentences, since children were better at rejecting sentences that did not share a meaning with original story information – semantically incongruent foils. Rather, the difficulty children exhibited distinguishing between original sentences and semantically congruent foils suggested that children were basing their recognition judgements on meanings of sentences, inferring that they had encoded meanings of the original material. Furthermore, the semantically congruent foils were constructed such that the information presented by each individual foil had been originally presented across two story sentences. The semantically congruent foils, therefore, tapped

whether the representations of story meaning encoded by children had been constructed by integrating information across sentences. In short, it was argued that false recognitions of semantically congruent foils reflected the integration of story information. Oakhill (1982) and Oakhill et al. (1986) found that less-skilled comprehenders made proportionately fewer false recognitions of semantically congruent foils than did the skilled comprehenders. It was concluded that the less-skilled comprehenders had been less likely than skilled comprehenders to integrate information across sentences.

However, in both of the Oakhill studies, (Oakhill, 1982; Oakhill et al., 1986) the recognition sentences were presented after the child had heard all eight stories and a further distracter activity. This procedure necessitated the processing of each story, while maintaining the representations of previous stories, and was therefore akin to complex working memory tasks, such as listening and reading span tasks used by Daneman and Carpenter (1980). Given the association between complex working memory and comprehension ability (see Sections 1.2.3 and 1.4.3), the storage demands of the original Oakhill paradigm might have exceeded the working memory capacities of less-skilled comprehenders, but not of skilled comprehenders. In short, the heavy memory demands of the sentence recognition test raise the possibility that less-skilled comprehenders may have exhibited reduced integrative representations of story information not because they failed to construct those representations, but because they failed to maintain them. This in turn might suggest that less-skilled comprehension of large pieces of text might not be due to a failure to comprehend individual sections of the text, but to a deficit in maintaining that comprehension across the remaining text.

Further findings that suggest that less-skilled comprehenders have a memory deficit rather than an integrative deficit are presented by Yuill, Oakhill, and Parkin (1989). Seven- and eight-year-old skilled and less-skilled comprehenders heard stories that contained an anomaly, such as a mother being pleased that a boy would not share his sweets with his brother. The information that resolved the anomaly (in the example case, the brother was on a diet) was

either adjacent to or at a distance from the anomaly. In order to account for the apparent anomaly, children needed to integrate the anomaly and the resolution information. Yuill et al. (1989) found no group difference in the adjacent condition, demonstrating that skilled and less-skilled comprehenders were equally able to integrate the two pieces of information. However, when the resolution was presented at a distance from the anomaly, less-skilled comprehenders were less able than skilled comprehenders to account for the anomaly. This pattern of results suggests that, whilst less-skilled comprehenders may have difficulty with memory for information across text, they are able to integrate that information when they have access to it.

This section has reviewed findings that could be considered evidence that less-skilled comprehenders exhibit reduced tendency to integrate information. However, the findings can also be accounted for by poor memory for information. In short, whilst Section 1.2.1 demonstrated that integration of information is a component process of comprehension, evidence that an integrative deficit underlies less-skilled comprehension is not yet conclusive.

1.4.2 Inference generation

By their very definition, inferences are what readers generate in order to fill the gaps in given information. It is, therefore, clear that reduced tendency to generate inferences would result in more informational gaps. Both the minimalist and constructionist views (see Section 1.2.2) agree that inferences are generated automatically in order to deconstruct the grammar of phrases, and to piece local information together, and that these inferences are necessary to comprehension. It therefore follows that reduced tendency to generate inferences might contribute to impaired comprehension. Indeed, some evidence exists to suggest that poor comprehenders and skilled comprehenders do differ in their degree of inference generation.

To examine the relationship between comprehension ability and tendency to generate inferences, Oakhill (1984) tested 7- and 8-year-old skilled and less-skilled comprehenders on their memory for information from read passages.

The questions tapped either literal information, that was expressed explicitly within the text, or inferential information, which the reader would have had to infer from textual information. The inferential information was deliberately what Oakhill (1984) referred to as implicit inferences, that is, inferences that were needed to make sense of the text, as opposed to explicit inferences which were deliberate deductions. Therefore, the inferences being tested corresponded to those needed for local coherence in the minimalist (McKoon & Ratcliff, 1992) and constructionist (Graesser et al., 1994) hypotheses for text comprehension, which both theories argue are essential for comprehension (see Section 1.2.2). Oakhill (1984) found that whilst skilled comprehenders answered generally more correct questions than less-skilled comprehenders from memory, both groups made more errors on inferential questions, to the same degree. This does not provide evidence that less-skilled comprehenders were particularly impaired at generating inferences; the data may have reflected poor memory in less-skilled comprehenders. However, when the text was made available for reference, skilled comprehenders made very few errors on both question types, and less-skilled comprehenders were very accurate on literal questions, but still made many errors on inferential questions. Oakhill (1984) claimed that this provided evidence that when controlling for memory, less-skilled comprehenders exhibited reduced tendency to generate inferences from read text. But there exists an alternative explanation for the results. It might be argued that the generation of inferences whilst reading a whole text might be qualitatively different from when scanning that text in order to answer a specific question, and it could be that ability to make a 'scanning' inference is dependent on familiarity with the text, both in layout and content. Since the less-skilled comprehension group in the work by Oakhill (1984) demonstrated poor memory of story information, their impairment in answering the inferential questions with reference to the text might not have been due to inability to make inferences, but in reduced access to text information from which to generate those inferences.

More persuasive evidence was provided by Long, Oppy, and Seely (1994), using a priming paradigm to investigate exactly which inferences were

automatically generated by skilled and less-skilled comprehenders during reading. Long et al. (1994) were interested in two types of inferences – those that were necessary to construct accurate discourse representation, i.e., inferences for meaning, and knowledge-based inferences that further embellished global meaning. Faster lexical decision to appropriate associate words (e.g., mint – candy) than inappropriate associate words would reveal the generation during reading of inferences after meaning. Faster decision to appropriate topic words (e.g., mint – breath) than inappropriate topic words would reveal the automatic generation of knowledge-based inferences. Long et al. (1994) found that skilled comprehenders responded faster to appropriate than inappropriate words of both inference type, indicating that they automatically generated both inferences for meaning and knowledge-based inferences during reading. But less-skilled comprehenders only responded faster to appropriate associate words – less-skilled comprehenders did not automatically generate knowledge-based inferences during reading. Furthermore, Long et al. (1994) assessed participants' knowledge of the topics used, and found no evidence that the less-skilled comprehenders' failure to generate knowledge-based inferences was due to poor knowledge.

Hannon and Daneman (1998) also found that less-skilled comprehenders were less likely to generate knowledge-based inferences. Although they did show evidence of generating such inferences when the text incorporated both a question that invited the inference and when the text was presented at a slower rate, neither of these aides were sufficient on their own. This suggests that while less-skilled comprehenders are capable of generating knowledge-based inferences during reading, they do not typically do so. Further evidence that less-skilled comprehenders do not automatically generate knowledge-based inferences lies in the finding that skilled comprehenders are better spontaneous instantiators (Oakhill, 1983). Instantiation is the substitution of a more specific interpretation of a word to suit the context in which it is used. For example, if told that a fish attacked a swimmer, adults will remember that the deed was done by a shark (Anderson & Ortony, 1975). This is an inference because it is an item of information built into the situation model that was never explicitly

stated in the text. Furthermore, it is a knowledge-base inference because it incorporates the comprehender's knowledge about the stereotypical behaviour of kinds of fish. Oakhill (1983) used original and instantiated word cues to prompt sentence recall, and found that skilled comprehenders' performance was better with instantiated cues than with original cues, indicating spontaneous instantiation. Less-skilled comprehenders' performance with both cue types, on the other hand, was equal to skilled comprehenders' performance with original cues, indicating that less-skilled comprehenders did not spontaneously instantiate. Furthermore, skilled comprehenders responded at the same speed to both cue types, indicating that instantiation had indeed taken place during text reading and not during recall, but less-skilled comprehenders – when they got the answer – took longer to respond to instantiated cues, indicating that when they did instantiate, they did so during recall and not during reading.

In summary, evidence suggests that whilst both skilled and less-skilled comprehenders generate inferences necessary to discourse processing, only skilled comprehenders generate knowledge-based, topic inferences. This evidence has two important implications. The first concerns the disagreement discussed in Section 1.2.2, regarding the type of inferences necessary to comprehension. The minimalists argue that global, topic inferences are not typically generated on-line during reading, and that they are not necessary for comprehension. However, Hannon and Daneman (1998), Long et al. (1994) and Oakhill (1983) demonstrated that generation of topic inferences and superior comprehension go hand in hand. This evidence, therefore, supports the constructionist hypothesis over the minimalist hypothesis. Secondly, the evidence strongly supports the notion that inference generation is an important component of comprehension, and that the success of individuals' comprehension might depend on the quality of inferences that they generate.

1.4.3 Memory

Section 1.2.3 reviewed the evidence that comprehension ability corresponds to working memory ability, investigating performance on simple span tasks that require simple storage of information, and complex span tasks that require

concurrent storage and processing. In brief summary, correlational studies found adult comprehension to not normally relate to performance on simple span tasks, but to relate to performance on complex span tasks, and child comprehension to relate to performance on both types of working memory task.

Commensurate with the low correspondence across the population between simple span and comprehension in adults, De Beni, Pallandino, Pazzaglia, and Cornoldi (1998) found that groups of young adults matched for logical reasoning but differing on a multiple choice reading comprehension test did not differ on a simple span task (digit recall). However, the between-group findings with children do not support the correlational data summarized above. Stothard and Hulme (1992) found that seven- and eight-year-old skilled and less-skilled comprehenders exhibited the same recall of digit strings, and Nation et al. (1999) found that ten-year-old skilled and less-skilled comprehenders did not differ on their ability to recall word lists. Furthermore, whilst phonological manipulations such as long versus short words, and words versus non-words are expected to reveal differences in phonological loop capacity, the comprehension groups investigated by Nation et al. (1999) were affected by phonological manipulations to the same degree. This evidence provides no support for the notion that working memory differences underlie comprehension deficits.

The correlational data revealed correspondence between complex working memory and comprehension in both adults and children. In line with this, De Beni et al. (1998) found that the groups of young adult skilled and less-skilled comprehenders mentioned above did differ on a complex working memory task, which required sentence verification with last word recall. This finding supports the evidence that low complex working memory ability accompanies less-skilled comprehension in adults.

Also in line with the correlational data is evidence that groups of children differing in comprehension ability also differ in complex working memory ability. The ten-year-old less-skilled comprehenders investigated by Nation et

al. (1999) exhibited lower listening spans (sentence veracity judgement plus last word recall) than skilled comprehenders. Similarly, the seven- and eight-year-old comprehension groups who participated in work conducted by Yuill et al. (1989) differed in a complex working memory task in which they read out digit strings and recalled the final digits. However, other studies have found no difference between comprehension groups on complex working memory tasks. Stothard and Hulme (1992) found no difference in listening span between seven- and eight-year-old skilled and less-skilled comprehenders, and Nation et al. (1999) found no spatial span difference between ten-year-old groups.

To summarise, De Beni et al. (1998) found that adult comprehension groups did not differ on a simple span task, but did differ on a complex span task. This pattern of results supports the correlational findings that adult comprehension is not related to simple storage ability but is related to ability to process and store information at the same time.

In children, the between-group findings do not present the same pattern as the correlational findings. Skilled and less-skilled comprehenders do not appear to differ in their simple storage ability (Nation et al., 1999; Stothard & Hulme, 1992), and while some evidence suggests that comprehension groups do differ on complex working memory tasks, other evidence refutes this finding. In short, the evidence does not conclusively support the notion that children's comprehension ability is related to their working memory or complex working memory ability.

1.5 Summary of introduction and overview of thesis

In this chapter, evidence was provided that integration of information, inference generation, and memory are component processes of comprehension, and that individuals with particularly poor comprehension ability exhibit reduced tendency to generate inferences. However, evidence concerning the relationship between poor comprehension and the other two processes – integration of information and memory – was not conclusive.

The experimental work presented in this thesis, therefore, examined whether integrative and memory deficiencies underlie poor comprehension ability. Further work explored the effects that different experimental conditions had on performance on comprehension assessments, and the relationship between performance on alternative assessments and a widely used comprehension test, the Neale Analysis of Reading Ability (Neale, 1958; Neale, 1997; Neale, Christophers, & Whetton, 1989).

CHAPTER 2

POOR COMPREHENSION, INTEGRATION AND MEMORY

In Section 1.1.2, the construction-integration theory of comprehension, developed by Kintsch (1988) and Kintsch and van Dijk (1978), was introduced. This theory proposes that information is integrated across individual propositions to construct a coherent mental model, and evidence suggests that this indeed takes place during successful comprehension (see Section 1.2.1).

The construction-integration model of comprehension also proposes that information is processed in cycles, each chunk being first maintained in working memory. It follows that the more information an individual can maintain in working memory, the fewer the number of cycles required, resulting in more efficient comprehension (see Section 1.2.3 for more detailed discussion). Evidence was reviewed that suggests some correspondence between working memory and comprehension abilities in the general population, particularly between comprehension and the ability to concurrently store and process information (see Section 1.2.3).

The experiments reported in this chapter investigate whether impairment in the ability to integrate or maintain information might lie at the root of specific comprehension difficulties. Such inquiries have been made before, but the evidence is not conclusive. Oakhill (1982) and Oakhill et al. (1986) administered a sentence recognition test, and found that less-skilled comprehenders were less likely than skilled comprehenders to falsely recognise sentences that they had not heard before but which integrated information across original sentences. It was claimed that this demonstrated an integrative deficit in the less-skilled comprehension group. However, since the procedure placed heavy memory demands on children, the group differences might alternatively reflect different effects of memory demands on the two groups (see Section 1.4.1 for detailed discussion). Section 1.4.3 reviewed the evidence that less-skilled comprehenders exhibit weaker complex working memory ability. Where such patterns of performance were found, the experimental

procedures made linguistic demands, such as sentence verification, or naming. The research reviewed in Section 1.3.1 demonstrates that less-skilled comprehenders tend to have semantic difficulties that would be expected to affect performance in such procedures, and therefore the weaker performance by less-skilled comprehenders in the complex working memory tasks reviewed in Section 1.4.3 might have arisen from difficulty with the processing aspect of the tasks as opposed to the memory aspect.

The experiments reported in this chapter therefore attempt to separate the integrative and memory demands of the sentence recognition test used by Oakhill (1982) and Oakhill et al. (1986), to investigate whether integrative deficits, or memory deficits, can account for less-skilled comprehension. In order to distinguish between integration of information and memory for information, the sentence recognition test was administered in two conditions that differed in memory demand. The blocked presentation condition, corresponding to Oakhill's original paradigm, was administered. This condition required children to listen to eight short stories in a block, and after a further delay testing for sentence recognition from all eight stories. This condition, therefore, required integration of information whilst also making heavy storage demands. The individual presentation condition, on the other hand, tested recognition of sentences immediately after presentation of an individual story. Whilst the processing demands remained the same – integration of information within one story at a time – children were no longer required to conduct this processing whilst maintaining previous story information. Performance on this version of the sentence recognition test, therefore, provided an index of the degree to which children had integrated information, without the representation of that information decaying due to complex working memory demands.

To summarise, the aims of the five experiments reported in this chapter were twofold. The first aim was to investigate whether less-skilled comprehenders were able to integrate semantic information within text when memory demands were minimised (Experiments 1 to 5). The second aim was to compare the effect of increased memory demand on the maintenance of semantic

information, in children with skilled and less-skilled reading comprehension (Experiments 2 to 5).

2.1 Experiment 1

In Experiment 1 the sentence recognition test was administered in the individual presentation condition, to investigate the tendency of skilled and less-skilled comprehenders to integrate information when memory demands were reduced. Findings of typical recognition performance by the children with less-skilled comprehension in this condition would represent a challenge to the view that poor integration of information underlies poor comprehension. The alternative finding – that less-skilled comprehenders falsely recognised fewer semantically congruent foils compared to skilled comprehenders, even with minimised memory load – would reinforce the view that poor comprehension is linked with poor integrative skills.

2.1.1 Method

Participants

One hundred and twenty-five children aged seven- and eight-years attending two urban primary schools in Bristol, south-west England, were screened on the Neale Analysis of Reading Ability (form 2) (Neale, 1997). This test provided age-related measures of children's ability to read aloud words in context and their comprehension of short passages. Thirty-two children were selected to participate in the experiment, according to profiles of performance on the Neale Analysis of Reading Ability. Sixteen less-skilled comprehenders were chosen, with at least age-appropriate accuracy, and comprehension age at least six months below reading accuracy age. Sixteen skilled comprehenders were selected, with at least age-appropriate reading accuracy and reading comprehension, and were selected to match the less skilled group for mean chronological age and accuracy.

The group profiles are summarised in Table 2.1. To verify the matching of these groups, a series of one-way analyses of variance, with group as factor,

UNIVERSITY
OF BRISTOL

were conducted on age, reading accuracy score and reading comprehension score. The analyses revealed non-significant main effects of group in age and accuracy, both $F(1,30)<1$, and significant main effect of group in comprehension, $F(1,30)=10.31$, $p<.01$.

Table 2.1. Participant characteristics in Experiment 1. Ages in years, standard deviations in parentheses.

	N			Neale Analysis of Reading Ability	
Comprehension group	Boys	Girls	Age	Accuracy age	Comprehension age
Skilled	11	5	8.0 (0.26)	9.2 (0.79)	8.6 (0.61)
Less-skilled	4	12	8.0 (0.27)	9.2 (0.75)	7.1 (1.80)

Materials

The stories and test sentences used by Oakhill (1982) and Oakhill et al. (1986) were employed in Experiment 1. The materials consisted of eight stories, each three sentences long. Each story was composed of two sentences describing an event (event sentence), and one describing a situation (stative sentence). The two event sentences had the same participant. Subject animacy was counterbalanced across stories. An example of the stories used is:

The man sat down behind the lady (event)
The lady was on the train (stative)
The man looked out of the window (event)

The recognition set for each story consisted of two sentences that were taken directly from the story and two new foil sentences. Of the foils, one was semantically congruent and the other semantically incongruent with the original story, but both were composed entirely from original vocabulary items. For example, the following four sentences comprised the recognition set for the above sample story:

The man sat down behind the lady (original sentence)
The lady was on the train (original sentence)

The man was on the train (semantically congruent foil)

The lady looked out of the window (semantically incongruent foil)

The original recognition sentences included one stative and one event sentence for each story. The two recognition foils were also composed such that there was one stative and one event foil for each story.

Design

Two-factor mixed design, with within-participant measure of recognition sentence type (16 original, 8 semantically congruent foils and 8 semantically incongruent foils), and between-participant measure of reading group (skilled comprehenders and less-skilled comprehenders). The dependent variable was number of sentences correctly recognised or rejected. All children received the same sentences. The order of stories and the order of recognition sentences within each set was randomised for each child.

Procedure

Children were familiar with the experimenter from the screening test sessions, and were tested individually in a quiet room. They were told that the experimenter was interested in children's memories for short stories. Children were instructed that they would hear one short story, which they had to listen to very carefully, and they would then hear some sentences that may or may not have been in the story, and they had to identify which sentences they had heard before. They were explicitly instructed that some 'trick' sentences might sound similar to something that had happened in the story, without having actually been in the story themselves. A practice trial was administered to acquaint children with the task. Each story and its accompanying set of test sentences was then presented one by one. A story was read aloud to the child, with a pause between each sentence. The experimenter then read out the four recognition sentences, to which children were instructed to answer "yes" if they had heard it before, and "no" if they had not heard it before. The experimenter recorded the children's responses at time of testing.

2.1.2 Results

Both reading groups were more accurate at recognising old sentences than at rejecting new ones. Both groups rejected more incongruent foils than congruent foils. Table 2.2 summarises correct recognitions and rejections of sentence types.

Table 2.2. Mean number of correct recognitions and rejections in Experiment 1. Standard deviations in parentheses.

	Recognitions	Rejections	
Comprehension group	Original sentences maximum16	Incongruent foils maximum 8	Congruent foils maximum 8
Skilled	13.3 (2.6)	6.4 (1.5)	4.5 (1.8)
Less-skilled	12.4 (2.0)	5.5 (1.4)	4.2 (1.7)

All analyses were conducted by participant, collapsing across materials (F_1), and by material, collapsing across participants (F_2).

The first stage of analysis examined whether the groups differed in their recognition accuracy. The one-way analyses of variance, with group as factor, were conducted on the number of correct responses. The analyses revealed non-significant main effect of group when analysed by participant, $F_1(1,30)=2.09$, $p>.05$, and significant main effect of group when analysed by item, $F_2(1,31)=6.23$, $p<.05$, with more correct responses provided by skilled comprehenders than less-skilled comprehenders.

The second stage of analysis examined whether groups were differentially affected by the semantic congruency of foils. Two-way analyses of variance were conducted on the number of foils rejected, with group and foil type as factor. The analyses revealed significant main effects of foil type, $F_1(1,30)=23.83$, $p<.001$, $F_2(1,14)=9.75$, $p<.01$, with more incongruent foils rejected than congruent foils, non-significant main effects of group, $F_1(1,30)=1.85$, $F_2(1,14)=4.04$, both $p>.05$, and non-significant interactions between group and foil type, $F_1(1,30)<1$, $F_2(1,14)=1.01$, $p>.05$.

In summary, both groups rejected more incongruent foils than congruent foils. Both groups were affected by semantic congruency of foils to the same extent.

2.1.3 Discussion

In Experiment 1, children who differed in text comprehension as measured by the Neale Analysis of Reading Ability (Neale, 1997) did not differ in their tendency to falsely recognise semantically congruent foils, relative to incongruent foils. This suggests that the comprehension groups did not differ in the degree to which they integrated story information.

Experiment 2 was designed firstly in order to determine whether the findings of Experiment 1 were replicable, and secondly to introduce a direct comparison of performance under different memory demands. The same two groups of skilled and less-skilled comprehenders were tested on their recognition of sentences, under the minimal memory demand of the individual presentation condition, and under the heavy memory demand of the blocked presentation condition – the version of the test administered by Oakhill and colleagues.

2.2 Experiment 2

2.2.1 Method

Participants

The same children who participated in Experiment 1 were tested again two weeks later.

Materials

Two new sets of eight three-line stories were constructed, each story being composed of two event and one stative sentence, with the two event sentences having the same subject. The recognition sentences for each story consisted of two original and two new sentences. The original sentences were drawn equally from the first and second or second and third positions in the original story to include one stative and one event sentence for each story. The two recognition

foils were also composed from the original story sentences so that there was one stative and one event sentence foil for each story. The full list of stories and recognition sentences used is presented in Appendix A.

Design

The design of Experiment 2 is the same as Experiment 1, with an additional within-participant measure of presentation condition (individual and blocked). Each child was presented one set of stories individually, and the other in a block. The orders of presentation condition, and the conditions in which each story-set was presented, were counterbalanced within each reading group. The order of stories and recognition sentences within the set were presented in a different random order for each participant. When the stories were presented as a block, the sets of recognition sentences were presented in the same order as the stories.

Procedure

Children were acquainted with the individual presentation condition from Experiment 1, and were told that they would be doing the familiar test again with different stories, and a slightly different version. They were reminded of the task, and the importance of not being 'tricked' by sentences that sounded like what happened in the story, but hadn't actually been in the story. In the individual presentation condition, each story followed by its accompanying test sentences was presented one by one, as in Experiment 1. In the blocked presentation condition, the children were told that they would listen very carefully to all of the stories, and then hear all of the test sentences. The stories were then read clearly to the child, with a slight pause between each sentence, and a longer pause between each story. The child and experimenter then played the card-matching game 'Snap' for three minutes as a distracter task. Then the recognition sentences were presented, and the child was instructed to say 'yes' if it had been heard before, and 'no' if not. The experimenter recorded responses at the time of testing.

2.2.2 Results

Both groups were more accurate at recognising old sentences than at rejecting new ones. Both groups rejected more semantically incongruent foils than congruent foils, in both presentation conditions. Table 2.3 summarises correct recognitions and rejections of sentence types.

Table 2.3. Mean number of correct recognitions and rejections in Experiment 2. Standard deviations in parentheses.

	Recognitions	Rejections	
	Original sentences	Incongruent foils	Congruent foils
Comprehension group	maximum 16	maximum 8	maximum 8
Following presentation of individual stories			
Skilled	13.6 (2.5)	7.1 (1.0)	5.1 (2.0)
Less-skilled	12.6 (1.7)	6.6 (1.0)	4.0 (1.5)
Following presentation of blocked stories			
Skilled	12.1 (2.6)	6.6 (1.7)	4.2 (1.9)
Less-skilled	12.3 (2.3)	6.1 (1.1)	4.2 (2.0)

The first stage of analysis investigated whether the groups differed in their recognition accuracy, in each condition separately. The one-way analyses of variance, with group as factor, conducted on the number of correct responses following individual presentation of stories revealed non-significant main effect of group when analysed by subject, $F_1(1,30)=3.79$, $p>.05$, and significant main effect of group when analysed by material, $F_2(1,63)=11.24$, $p=.001$, with more correct answers provided by skilled comprehenders than by less-skilled comprehenders. The one-way analyses of variance, with group as factor, conducted on accuracy following blocked presentation of stories revealed non-significant main effects of group, both $F<1$.

To investigate whether presentation condition differentially affected the accuracy of the groups, two-way analyses of variance, with group and presentation condition as factors, were conducted on the number of correct responses. The analyses revealed significant main effects of condition,

$F_1(1,30)=4.83$, $F_2(1,63)=9.61$, both $p<.05$, with more correct responses provided after stories were presented individually than after stories were presented in a block. The analyses revealed non-significant main effect of group when analysed by subject, $F_1(1,30)=1.43$, $p>.05$, and significant main effect of group when analysed by material, $F_2(1,63)=7.98$, $p<.01$, with more correct responses provided by skilled comprehenders than by less-skilled comprehenders. The analyses revealed non-significant interactions between condition and group, $F_1(1,30)=1.90$, $F_2(1,63)=3.36$, both $p>.05$.

The second stage of analysis investigated whether the groups were differentially affected by the congruency of foils, in each condition separately. Two-way analyses of variance, with group and foil type as factors, were conducted on the number of foils rejected following individual presentation of stories. The analyses revealed significant main effects of foil type, $F_1(1,30)=73.54$, $F_2(1,30)=54.79$, both $p<.001$, with more incongruent foils rejected than congruent foils. The analyses revealed non-significant main effect of group when analysed by subject, $F_1(1,30)=3.82$, $p>.05$, and significant main effect of group when analysed by item, $F_2(1,30)=9.03$, $p<.01$, with more foils rejected by skilled comprehenders than by less-skilled comprehenders. The analyses of variance revealed non-significant interactions between group and foil types, $F_1(1,30)=1.12$, $F_2(1,30)=1.00$, both $p>.05$.

Two-way analyses of variance were then conducted, with group and foil type as factors, on the number of foils rejected following blocked presentation of stories. The analyses revealed significant main effects of foil type, $F_1(1,30)=36.28$, $F_2(1,30)=32.47$, both $p<.001$, with more incongruent foils rejected than congruent foils, non-significant main effects of group, both $F<1$, and non-significant interactions between group and foil type, both $F<1$.

Finally, to investigate whether presentation condition differentially affected group rejection of foils, three-way analyses of variance, with condition, group and foil type as factors, were conducted on the numbers of foils rejected. The analyses revealed significant main effects of foil type, $F_1(1,30)=93.34$,

$F_2(1,30)=58.72$, both $p<.001$, with more incongruent foils rejected than congruent foils, non-significant main effect of test condition when analysed by subject, $F_1(1,30)=3.35$, $p>.05$, and significant main effect of test condition when analysed by material, $F_2(1,30)=5.98$, $p<.05$, with more foils rejected after stories were presented individually than as a block. The analyses also revealed non-significant main effect of group when analysed by subject, $F_1(1,30)=1.91$, $p>.05$, and significant main effect of group when analysed by material, $F_2(1,30)=7.33$, $p<.05$, with more foils rejected by skilled comprehenders than by less-skilled comprehenders. The analyses found non-significant interactions between group and foil type, both $F<1$, between group and condition, $F_1(1,30)=1.44$, $F_2(1,30)=1.58$, both $p>.05$, between condition and foil type, both $F<1$, and between group, condition and foil type, $F_1(1,30)=1.55$, $F_2(1,30)=1.26$, both $p>.05$.

Both groups made more accurate recognitions and rejections after stories were presented individually than after they were presented in a block. Both groups correctly rejected more incongruent foils than congruent foils. Furthermore, less-skilled comprehenders were affected by semantic congruency of foils to the same degree as the skilled comprehenders, in both presentation conditions.

2.2.3 Discussion

Experiment 2 replicated the finding of Experiment 1 that skilled and less-skilled comprehenders both found it more difficult to reject foils that were congruent with story meaning than foils that were incongruent, following presentation of individual stories. This evidence suggests that both skilled and less-skilled comprehenders based their recognition judgements on story meaning. Furthermore, since the congruent foils presented information that had been combined across original sentences, the findings indicate that both skilled and less-skilled comprehenders integrated information across sentences when encoding story meaning.

Overall recognition accuracy was better after hearing one story than after hearing a number of stories, demonstrating that accuracy decreased under

increased memory demand. Furthermore, less-skilled comprehenders were affected by increased memory demands between the two conditions to the same degree as skilled comprehenders. This finding is contrary to the evidence that suggests that individuals differing in comprehension ability differ in performance on tasks that require concurrent processing and storage of linguistic materials (see Sections 1.2.3 and 1.4.3).

In addition, the comprehension groups were affected by semantic congruency of foil to the same extent in the blocked condition of Experiment 2. This finding is inconsistent with those reported by Oakhill (1982) and Oakhill et al. (1986), from an identical procedure. However, a difference exists between the materials used by Oakhill and colleagues, and those used in Experiment 2, in that the latter were not balanced for animacy. In order to verify that the surprising results of the blocked condition of Experiment 2 were not an artifact of this imbalance, the experiment was repeated, with new materials, in which the subjects of half the stories were animate, and half inanimate. In Experiment 3, the same children were again tested for their recognition of original sentences and rejection of semantically congruent and incongruent foils, after stories were presented either individually, or in a block. New stories were constructed for the purposes of this experiment in order to test the generality of the findings in Experiments 1 and 2.

2.3 Experiment 3

2.3.1 Method

Participants

The same children who participated in Experiments 1 and 2 were tested again, six weeks after Experiment 2.

Materials

Two new sets of eight three-line stories were composed in the manner described in Experiment 2, with the additional constraint that the subjects of

half the stories in each set were animate, and the other half inanimate. The full list of stories and recognition sentences used is presented in Appendix B.

Design and Procedure

The design and procedure were the same as described in Experiment 2.

2.3.2 Results

Both groups were more accurate at recognising old sentences than at rejecting new ones. Both groups rejected more semantically incongruent foils than congruent foils, in both presentation conditions. Table 2.4 summarises correct recognitions and rejections of sentence types.

Table 2.4. Mean number of correct recognitions and rejections in Experiment 3. Standard deviations in parentheses.

	Recognitions	Rejections	
	Original sentences	Incongruent foils	Congruent foils
Comprehension group	maximum 16	maximum 8	maximum 8
Following presentation of individual stories			
Skilled	12.9 (2.0)	7.1 (1.2)	4.3 (2.0)
Less-skilled	12.6 (1.7)	6.6 (1.1)	4.5 (1.6)
Following presentation of blocked stories			
Skilled	11.1 (2.8)	6.3 (1.5)	4.3 (1.5)
Less-skilled	11.3 (2.1)	6.2 (1.3)	4.6 (1.2)

The first stage of analysis investigated whether the groups differed in their recognition accuracy. One-way analyses of variance, with group as factor, were conducted on the number of correct responses in each presentation condition separately. The analyses revealed non-significant main effects of group following individual presentation of stories, $F_1(1,30)<1$, $F_2(1,63)=1.17$, $p>.05$, and following blocked presentation of stories, both $F<1$.

To investigate whether presentation condition differentially affected the accuracy of the groups, two-way analyses of variance, with group and

presentation condition as factors, were conducted on the number of correct responses. The analyses revealed significant main effects of condition, $\underline{F}_1(1,30)=7.92$, $\underline{F}_2(1,63)=7.23$, both $\underline{p}<.01$, with more correct responses provided after stories were presented individually than after stories were presented in a block, non-significant main effects of group, both $\underline{F}<1$, and non-significant interactions between condition and group, $\underline{F}_1(1,30)<1$, $\underline{F}_2(1,63)=1.14$, $\underline{p}>.05$.

The second stage of analysis investigated whether the groups were differentially affected by the congruency of foils. Two-way analyses of variance, with group and foil type as factors, were conducted on the number of foils rejected. The analyses revealed significant main effects of foil type, $\underline{F}_1(1,30)=72.10$, $\underline{F}_2(1,30)=21.09$, both $\underline{p}<.001$, with more incongruent foils rejected than congruent foils, non-significant main effects of group, both $\underline{F}<1$, and non-significant interactions between group and foil type, $\underline{F}_1(1,30)=1.42$, $\underline{F}_2(1,30)=2.55$, both $\underline{p}>.05$.

Two-way analyses of variance were then conducted, with group and foil type as factors, on the number of foils rejected following blocked presentation of stories. The analyses revealed significant main effects of foil type, $\underline{F}_1(1,30)=35.14$, $\underline{F}_2(1,30)=18.38$, both $\underline{p}<.001$, with more incongruent foils rejected than congruent foils, non-significant main effects of group, both $\underline{F}<1$, and non-significant interactions between group and foil type, both $\underline{F}<1$.

Finally, to investigate whether presentation condition differentially affected group rejection of foils, three-way analyses of variance, with condition, group and foil type as factors, were conducted on the numbers of foils rejected. The analyses revealed significant main effects of foil type, $\underline{F}_1(1,30)=81.79$, $\underline{F}_2(1,30)=32.12$, both $\underline{p}<.001$, with more incongruent foils rejected than congruent foils, non-significant main effects of test condition, $\underline{F}_1(1,30)=1.47$, $\underline{F}_2(1,30)=1.08$, both $\underline{p}>.05$, and non-significant main effects of group, both $\underline{F}<1$. The analyses also found non-significant interactions between group and foil type, $\underline{F}_1(1,30)=1.13$, $\underline{F}_2(1,30)=1.79$, both $\underline{p}>.05$, between condition and foil

type, $\underline{F}_1(1,30)=3.73$, $\underline{F}_2(1,30)=1.31$, both $\underline{p}>.05$, between condition and group, both $\underline{F}<1$, and between group, condition and foil type, both $\underline{F}<1$.

Both groups made more accurate recognitions and rejections after stories were presented individually than after they were presented in a block. Both groups correctly rejected more incongruent foils than congruent foils. Furthermore, less-skilled comprehenders were affected by semantic congruency of foils to the same degree as the skilled comprehenders, in both presentation conditions.

2.3.3 Discussion

In Experiment 3, sentence recognition was more accurate after presentation of individual stories than after presentation of a block of stories. In addition, less-skilled comprehenders were affected by increased memory demand to the same degree as skilled comprehenders. This finding is not well accommodated by the suggestion that less-skilled comprehenders exhibit impaired complex working memory capacity compared to skilled comprehenders (see Section 1.4.3).

Experiment 3 replicated the finding of Experiments 1 and 2 that skilled and less-skilled comprehenders did not differ in their tendency to integrate information under reduced memory demands. Furthermore, the groups did not differ in their ability to maintain and recognise that integrated information, replicating the finding of Experiment 2. This pattern of results in the blocked presentation condition of Experiments 2 and 3 is contrary to those found by Oakhill (1982) and Oakhill et al. (1986) from an identical task.

This failure to replicate Oakhill and colleagues might reflect the effect of practice in the present research. All children were introduced to the sentence recognition test with recognition sentences presented immediately after individual stories. They may have become aware of the manner in which the foils were constructed, thereby affecting their subsequent performance in the blocked condition of Experiments 2 and 3. Oakhill and colleagues, on the other hand, administered only the blocked condition of the sentence recognition test.

Experiment 4, therefore, investigated the effect of the individual condition on performance in the blocked condition, by administering both conditions of the sentence recognition test to new groups of skilled and less-skilled comprehenders. The order in which the two conditions were presented was counterbalanced within groups. An effect of order would have suggested that the failure to replicate the findings of Oakhill (1982) and Oakhill et al. (1986) in Experiments 2 and 3 might be accounted for by the difference in test administration. Thus, Experiment 4 sought to examine a potential account for the discrepancy between Experiments 2 and 3 and findings by Oakhill and colleagues, with new groups of skilled and less-skilled comprehenders.

2.4 Experiment 4

2.4.1 Method

Participants

Ninety-seven children aged seven- and eight-years attending two urban primary schools in Bristol were screened on the Neale Analysis of Reading Ability (form 1) (Neale, 1997). Sixteen less-skilled comprehenders and sixteen skilled comprehenders were selected to participate in the experiment, according to the criteria specified in Experiment 1.

The group profiles are summarised in Table 2.5. To verify the matching of these groups, a series of two-way analyses of variance were conducted, by participant, with group and order as factors, on age, accuracy score and comprehension score separately. The analysis of age revealed non-significant main effects of group and order, and non-significant interaction between group and order, all $F_1(1,28)<1$. The analysis of accuracy score also revealed non-significant main effects of group, $F_1(1,28)=1.74$, $p>.05$, and order, $F_1(1,28)=1.53$, both $p>.05$, and non-significant interaction between group and order, $F_1(1,28)<1$. The analysis of comprehension score revealed significant main effect of group, $F_1(1,28)=9.56$, $p<.01$, with skilled comprehenders having

higher scores than less-skilled comprehenders, non-significant main effect of order and non-significant interaction between group and order, both $F_1(1,28)<1$.

Table 2.5. Participant characteristics for Experiment 4, by order of presentation condition. Ages in years, standard deviations in parentheses.

	N			Neale Analysis of Reading Ability	
Comprehension group	Boys	Girls	Age	Accuracy age	Comprehension age
Skilled					
Individual condition first	4	4	8.1 (0.24)	9.2 (1.3)	9.0 (1.6)
Blocked condition first	5	3	8.2 (0.30)	8.6 (1.2)	8.5 (1.0)
Less-skilled					
Individual condition first	2	6	8.1 (0.32)	9.8 (1.6)	7.7 (1.0)
Blocked condition first	3	5	8.1 (0.21)	9.3 (1.2)	7.7 (0.56)

Materials

As Experiment 3

Design

As Experiment 2, with additional factor of presentation order - individual condition first, or blocked condition first.

Procedure

As Experiment 2.

2.4.2 Results

Both groups were more accurate at recognising old sentences than at rejecting new sentences, both when stories were presented individually and in a block. Both groups rejected more incongruent foils than congruent foils, in both presentation conditions. Table 2.6 summarises correct recognitions and rejections of sentence types.

Initial analyses were carried out on data combined across order of presentation.

Table 2.6. Mean number of correct recognitions and rejections in Experiment 4, by order of presentation condition. Standard deviations in parentheses.

	Recognitions	Rejections	
Comprehension group	Original sentences maximum 16	Incongruent foils maximum 8	Congruent foils maximum 8
Following presentation of individual stories			
Skilled			
Individual condition first	14.2 (2.0)	6.5 (1.8)	3.5 (1.5)
Blocked condition first	14.1 (1.0)	7.1 (1.1)	3.5 (3.0)
Less-skilled			
Individual condition first	13.8 (2.1)	6.0 (1.9)	3.1 (1.4)
Blocked condition first	13.1 (2.0)	6.6 (1.1)	3.1 (1.1)
Following presentation of blocked stories			
Skilled			
Individual condition first	14.0 (1.9)	6.4 (1.1)	5.8 (2.0)
Blocked condition first	13.5 (2.2)	5.4 (2.7)	2.8 (2.1)
Less-skilled			
Individual condition first	11.1 (1.6)	5.3 (1.8)	2.8 (1.8)
Blocked condition first	12.1 (1.1)	5.4 (1.6)	3.1 (1.7)

The first stage of analysis investigated whether the groups differed in their recognition accuracy. One-way analyses of variance, with group as factor, were conducted on number of correct responses following individual presentation of stories. The analyses revealed non-significant main effects of group, $F_1(1,30)=2.54$, $F_2(1,63)=3.51$, both $p>.05$. One-way analyses of variance, with group as factor, were then conducted on accuracy following blocked presentation of stories. The analyses revealed non-significant main effect of group when analysed by participant, $F_1(1,30)=3.33$, $p>.05$, and significant main effect of group when analysed by item, $F_2(1,63)=7.75$, $p<.01$, with more correct responses provided by skilled comprehenders than less-skilled comprehenders.

To investigate whether presentation condition differentially affected group accuracy, two-way analyses of variance were conducted on the number of

correct responses, with group and presentation condition as factors. The analyses revealed significant main effects of condition, $F_1(1,30)=9.03$, $p<.01$, $F_2(1,63)=12.36$, $p=.001$, with more correct responses provided after stories were presented individually than after stories were presented in a block, and significant main effects of group, $F_1(1,30)=6.62$, $p<.05$, $F_2(1,63)=10.29$, $p<.01$, with more correct responses provided by skilled comprehenders than by less-skilled comprehenders, and non-significant interactions between condition and group, both $F<1$.

The next stage of analysis investigated whether the groups were differentially affected by the congruency of foils. Two-way analyses of variance, with group and foil type as factors, were conducted on the number of foils rejected following individual presentation of stories. The analyses revealed significant main effects of foil type, $F_1(1,30)=62.64$, $F_2(1,30)=72.69$, both $p<.001$, with more incongruent foils rejected than congruent foils, non-significant main effects of group, $F_1(1,30)=1.09$, $F_2(1,30)=1.71$, both $p>.05$, and non-significant interactions between foil and group, both $F<1$.

Next, two-way analyses of variance, with group and foil type as factors, were conducted on the number of foils rejected following blocked presentation of stories. The analyses revealed significant main effects of foil type, $F_1(1,30)=94.04$, $F_2(1,30)=144.46$, both $p<.001$, with more incongruent foils rejected than congruent foils, non-significant main effects of group, both $F<1$, and non-significant interactions between group and foil type, $F_1(1,30)=2.28$, $F_2(1,30)=1.81$, both $p>.05$.

To investigate whether presentation condition differentially affected foil rejections, three-way analyses of variance, with condition, group and foil type as factors, were conducted on the number of foils rejected. The analyses revealed significant main effects of foil type, $F_1(1,30)=166.53$, $F_2(1,30)=49.57$, both $p<.001$, with more incongruent foils rejected than congruent foils, and significant main effects of test condition, $F_1(1,30)=6.41$, $F_2(1,30)=6.36$, both $p<.05$, with more foils rejected after stories were presented individually than in

a block. The analyses revealed non-significant main effects of group, $F_1(1,30)<1$, $F_2(1,30)=1.62$, $p>.05$, non-significant interactions between group and foil type, $F_1(1,30)=1.13$, $p>.05$, $F_2(1,30)=1.28$, $p>.05$, and non-significant interactions between condition and foil type, between condition and group, and between condition, group and foil type, all $F<1$.

The final analyses investigated whether the order in which the conditions were presented affected group rejection of foils following blocked presentation of stories. Three-way analyses of variance, with group, foil type and presentation order as factors, were conducted on the number of foils rejected in the blocked condition. The analyses revealed significant main effects of foil type, $F_1(1,28)=95.46$, $F_2(1,30)=19.65$, both $p<.001$, with more semantically incongruent foils rejected than congruent foils. The analyses revealed non-significant main effects of presentation order, both $F<1$, and non-significant interactions between order and group, both $F<1$, between order and foil type, $F_1(1,28)=1.70$, $F_2(1,30)=1.31$, both $p>.05$, and between order, group and foil type, both $F<1$.

In summary, both groups were more accurate after stories had been presented individually than after they had been presented in a block. Both groups rejected more semantically incongruent foils than congruent foils, and the groups were affected by semantic congruency of foils to the same degree. Furthermore, the extent to which groups were affected by the semantic congruency of foils was not affected by the order in which they received the presentation conditions.

2.4.3 Discussion

Experiment 4 replicated the findings of Experiments 1, 2 and 3 that skilled and less-skilled comprehenders exhibited the same tendency to integrate information from heard stories, with new groups. In addition, Experiment 4 replicated the finding of Experiments 2 and 3 that recognition accuracy decreased with increased memory demand, and that skilled and less-skilled comprehenders were affected by increased memory demand to the same degree.

Furthermore, the above pattern of results remained consistent in Experiment 4 across order of presentation. The children who participated in Experiments 2 and 3 had been introduced to the task in the less demanding individual presentation paradigm (in Experiment 1), but the children who participated in the experiments conducted by Oakhill (1982) and Oakhill et al. (1986) had been introduced to the task in the more demanding blocked presentation paradigm. In Experiment 4, the condition in which the task was presented did not affect performance. Therefore, the discrepancy between the current findings and the findings of Oakhill and colleagues do not appear to be accounted for by the differences in experimental administration.

One remaining concern is that the skilled and less-skilled comprehension groups of the current work might not have differed sufficiently on comprehension ability. The selection criteria defined less-skilled comprehenders as children with at least age appropriate reading accuracy, and reading comprehension at least six months behind their accuracy. This would enable a child with good accuracy (six months or more ahead of that expected by age) and average comprehension to be categorized as a less-skilled comprehender. Certainly then, a group of such children would not be expected to reflect cognitive profiles that underlie genuine poor comprehension.

Experiment 5 was conducted in order to investigate differences between skilled and less-skilled comprehenders on the two conditions of the sentence recognition test, with a group of children with more extreme comprehension impairments. The comprehension deficit in the selection criteria was increased to at least a year below that expected from age. This extended the margin between the groups, beyond individual fluctuation in performance, and avoided the inclusion of children with very good reading accuracy and average comprehension in the less-skilled comprehension group. In addition, complex working memory assessments were included in order to provide explicit working memory comparisons between the groups. In short, Experiment 5 was conducted to verify the findings of Experiments 1 to 4, or the findings of Oakhill and colleagues, with a new group of less-skilled comprehenders whose

UNIVERSITY OF BRISTOL

comprehension impairment was more severe than groups previously investigated in the present work and by Oakhill and colleagues.

2.5 Experiment 5

2.5.1 Method

Participants

One hundred and seventy five children aged seven- and eight-years attending five urban primary schools in Bristol were screened on the Neale Analysis of Reading Ability (form 1) (Neale, 1997). Sixteen skilled comprehenders were selected to participate, with (a) at least age-appropriate reading accuracy and comprehension, (b) reading accuracy and comprehension ages no more than two years above actual age, and (c) reading accuracy and comprehension ages within 12 months of each other. Sixteen less-skilled comprehenders were selected, who exhibited (a) at least age-appropriate reading accuracy, and (b) reading comprehension age at least 12 months below actual age. Children were also screened on the British Picture Vocabulary Scale (Dunn, Dunn, Whetton, & Burley, 1997), and only children who achieved standardised score within one standard deviation from the mean were included. This controlled for the possible effect of vocabulary on performance on the Neale Analysis of Reading Ability. Skilled comprehenders obtained mean vocabulary age of 8.00 years (standard deviation 0.62 years), and less-skilled comprehenders obtained mean vocabulary age of 7.94 years (standard deviation 0.70 years).

The group profiles are summarised in Table 2.7. To verify the matching of these groups, a series of one-way analyses of variance were conducted, with group as factor, on age, vocabulary score, accuracy score and comprehension score separately. These revealed non-significant main effects of group on age, $F_1(1,30)<1$, on vocabulary, $F_1(1,30)<1$, and on accuracy, $F_1(1,30)=3.25$, $p>.05$, and significant main effect of group on comprehension, $F_1(1,30)=185.67$, $p<.001$.

Table 2.7. Participant characteristics in Experiment 5, by order of presentation condition. Ages in years, standard deviations in parentheses.

	N			Neale Analysis of Reading Ability	
Comprehension group	Boys	Girls	Age	Accuracy age	Comprehension age
Skilled					
Individual condition first	3	5	8.9 (0.23)	8.5 (0.29)	8.5 (0.46)
Blocked condition first	2	6	9.0 (0.26)	8.4 (0.58)	8.6 (0.50)
Less-skilled					
Individual condition first	4	4	8.9 (0.26)	8.8 (1.1)	6.5 (0.46)
Blocked condition first	3	5	8.9 (0.38)	9.0 (0.89)	6.4 (0.39)

Materials

As Experiment 3. The counting span and listening span tests of the Working Memory Test Battery (Pickering & Gathercole, 2001) were also administered, in addition to the British Picture Vocabulary Scale (Dunn et al., 1997).

Design

As Experiment 4, with two additional between-participant measures of listening span and counting span.

Procedure

The administration of the sentence recognition test was as described in Experiment 2. The counting span was administered in the following way. Children were instructed to count aloud the number of dots on a series of cards, and to then recall the number of dots on each card, in the order that they were seen. Responses were recorded at testing. The test contained six examples of each list-size. Children continued through increasing list-lengths until they correctly recalled fewer than four out of six lists of a list-length. Performance was scored as the average number of items on the last six trials correctly recalled in the correct order. Children were then administered the listening span test. They were told that they would listen to a series of sentences, and were instructed to say whether each sentence was true or false, and then to recall the final word of each sentence in the order that they had been heard. Responses

were noted at testing. The listening span test was structured and scored in the same manner as the counting span test.

2.5.2 Results

Both groups were more accurate at recognising old sentences than at rejecting new sentences, both when stories were presented individually and in a block. Both groups rejected more incongruent foils than congruent foils, in both presentation conditions. Table 2.8 summarises correct recognitions and rejections of sentence types.

Table 2.8. Mean number of correct recognitions and rejections in Experiment 5, by order of presentation condition. Standard deviations in parentheses.

	Recognitions	Rejections	
Comprehension group	Original sentences maximum 16	Incongruent foils maximum 8	Congruent foils maximum 8
Following presentation of individual stories			
Skilled			
Individual condition first	13.9 (1.6)	6.0 (2.1)	3.4 (1.5)
Blocked condition first	14.3 (1.6)	7.6 (0.5)	3.8 (1.3)
Less-skilled			
Individual condition first	13.8 (1.4)	6.0 (1.4)	3.0 (1.2)
Blocked condition first	13.6 (2.1)	5.5 (2.5)	2.5 (2.3)
Following presentation of blocked stories			
Skilled			
Individual condition first	11.8 (2.8)	5.5 (1.3)	2.4 (1.2)
Blocked condition first	13.0 (2.0)	6.3 (0.89)	3.6 (1.6)
Less-skilled			
Individual condition first	11.1 (2.4)	5.1 (1.6)	2.5 (2.1)
Blocked condition first	10.3 (3.0)	5.4 (1.8)	2.9 (2.0)

Initial analyses were carried out on data combined across order of presentation.

The first stage of analysis investigated whether the groups differed in recognition accuracy. One-way analyses of variance, with group as factor, were conducted on the number of correct responses following individual presentation of stories. The analyses revealed significant main effects of group, $F_1(1,30)=4.81$, $p<.05$, $F_2(1,63)=7.43$, $p<.01$, with more correct responses provided by skilled comprehenders than by less-skilled comprehenders. Next, one-way analyses of variance, with group as factor, were conducted on accuracy following blocked presentation of stories. The analyses revealed non-significant main effect of group when analysed by participant, $F_1(1,30)=3.67$, $p>.05$, and significant main effect of group when analysed by item, $F_2(1,63)=9.02$, $p<.01$, with more correct responses provided by skilled comprehenders than by less-skilled comprehenders.

To investigate whether presentation condition differentially affected the accuracy of the groups, two-way analyses of variance were conducted on the number of correct responses, with group and presentation condition as factors. The analyses revealed significant main effects of condition, $F_1(1,30)=20.77$, $F_2(1,63)=29.01$, both $p<.001$, with more correct responses after stories were presented individually than as a block, significant main effects of group, $F_1(1,30)=6.48$, $p<.05$, $F_2(1,63)=15.16$, $p<.001$, with more correct responses provided by skilled comprehenders than by less-skilled comprehenders, and non-significant interactions between condition and group, both $F<1$.

The second stage of analysis investigated whether the groups were differentially affected by the congruency of foils. Two-way analyses of variance, with group and foil type as factors, were conducted on the number of foils rejected following individual presentation of stories. The analyses revealed significant main effects of foil type, $F_1(1,30)=80.13$, $F_2(1,30)=54.71$, both $p<.001$, with more incongruent foils rejected than congruent foils. The analysis also revealed non-significant main effect of group when analysed by participant, $F_1(1,30)=3.64$, $p>.05$, and significant main effect of group when analysed by item, $F_2(1,30)=7.12$, $p<.05$, with more foils rejected by skilled

comprehenders than by less-skilled comprehenders. The analyses revealed non-significant interactions between foil type and group, both $\underline{F}<1$.

Next, two-way analyses of variance, with group and foil type as factors, were conducted on the number of foils rejected following blocked presentation of stories. The analyses revealed significant main effects of foil type, $\underline{F}_1(1,30)=53.08$, $\underline{F}_2(1,30)=32.98$, both $\underline{p}<.001$, with more incongruent foils rejected than congruent foils, non-significant main effects of group, $\underline{F}_1(1,30)=1.21$, $\underline{F}_2(1,30)=2.17$, both $\underline{p}>.05$, and non-significant interactions between group and foil type, both $\underline{F}<1$.

To investigate whether presentation condition differentially affected foil rejection, three-way analyses of variance, with condition, group and foil type as factors, were conducted on the number of foils rejected. The analyses revealed significant main effects of foil type, $\underline{F}_1(1,30)=94.84$, $\underline{F}_2(1,30)=62.39$, both $\underline{p}<.001$, with more incongruent foils rejected than congruent foils. The analyses revealed non-significant main effect of test condition when analysed by participant, $\underline{F}_1(1,30)=3.45$, $\underline{p}>.05$, and near-significant main effect of condition when analysed by item, $\underline{F}_2(1,30)=4.12$, $\underline{p}=.051$. The analyses also revealed non-significant main effect of group when analysed by participant, $\underline{F}_1(1,30)=3.67$, $\underline{p}>.05$, and significant main effect of group when analysed by item, $\underline{F}_2(1,30)=7.74$, $\underline{p}<.01$, with more foils rejected by skilled comprehenders than by less-skilled comprehenders. The analyses revealed non-significant interactions between group and foil type, both $\underline{F}<1$, between condition and foil type, $\underline{F}_1(1,30)=1.02$, $\underline{p}>.05$, $\underline{F}_2(1,30)<1$, between condition and group, $\underline{F}_1(1,30)<1$, $\underline{F}_2(1,30)=1.14$, $\underline{p}>.05$, and between condition, group and foil type, both $\underline{F}<1$.

The next analyses investigated whether order of condition affected group performance in the blocked presentation condition. Three-way analyses of variance, with group, foil type and presentation order as factors, were conducted on the number of foils rejected following blocked presentation. These revealed significant main effects of foil type, $\underline{F}_1(1,28)=49.94$,

$F_2(1,30)=32.98$, both $p<.001$, with more incongruent foils rejected than congruent foils, non-significant main effects of presentation order, $F_1(1,28)=2.46$, $F_2(1,30)=3.74$, both $p>.05$, non-significant main effects of group, $F_1(1,28)=1.25$, $F_2(1,30)=2.17$, both $p>.05$, and non-significant interactions between order and group, $F_1(1,28)<1$, $F_2(1,30)=1.00$, $p>.05$. The analyses also revealed non-significant interactions between order and foil type, between order, group and foil type, and between group and foil type, all $F<1$.

The final analyses examined whether the groups differed in their performance on the complex working memory tasks. Skilled comprehenders obtained mean counting span of 2.27 (standard deviation = 0.91), and mean listening span of 1.70 (standard deviation = 0.45). Less-skilled comprehenders obtained mean counting span of 2.82 (standard deviation = 0.97), and mean listening span of 1.82 (standard deviation = 0.49). One-way analyses of variance, by participant, with group as factor, were conducted on the counting and listening spans scores separately, and revealed non-significant main effects of group on counting span, and on listening span, both $F_1(1,30)<1$.

In summary, both groups made more accurate recognitions and rejections of sentences after stories had been presented individually than after they had been presented in a block. Both groups rejected more semantically incongruent foils than congruent foils. Although less-skilled comprehenders made generally fewer correct judgments, the groups were affected by semantic congruency of foils to the same degree. Furthermore, the extent to which groups were affected by the semantic congruency of foils was not affected by the order in which they received the presentation conditions. The comprehension groups did not differ in their performance on listening and counting complex working memory tasks, nor in the extent to which their processing and maintenance of story information was affected by increased memory load.

2.5.3 Discussion

Experiment 5 replicated the findings of Experiments 1 to 4 that skilled and less-skilled comprehenders exhibited the same tendency to integrate information

from heard stories, and the findings of Experiments 2 and 4 that recognition accuracy decreased with increased memory demand, and that skilled and less-skilled comprehenders were affected by increased demand to the same degree. Experiment 5 also replicated the finding of Experiment 4 that the presentation condition in which the sentence recognition test was introduced did not affect performance on the test.

One possible explanation for the discrepancy between the findings of the current work and those of Oakhill and colleagues was that the comprehension groups participating in the current work did not differ sufficiently in comprehension ability. This suspicion might be heightened by the evidence that the comprehension groups who participated in Experiment 5 did not differ in their performance on two complex working memory tasks and were affected by increased memory demand in the sentence recognition test to the same degree. These findings are contrary to existing evidence demonstrating correspondence between comprehension ability and performance on complex working memory tasks with linguistic material (see Sections 1.2.3 and 1.4.3). However, this explanation seems unlikely, since the comprehension deficit of the less-skilled comprehension group who participated in Experiment 5 was quite severe. Furthermore, group differences did exist in the current work, with less-skilled comprehenders exhibiting generally lower recognition accuracy in Experiment 5 than skilled comprehenders.

2.6 General Discussion

The work presented in this chapter set out to investigate whether poor integrative abilities or poor memory might underlie low comprehension ability. Each aspect of the investigation will be discussed in turn. Firstly, Experiments 1 to 5 provided evidence that, when the memory demands of a sentence recognition test were minimised, less-skilled comprehenders were just as likely as skilled comprehenders to integrate information in aurally presented stories. This suggests that poor comprehenders do not suffer from an integrative deficit, and that the previous evidence to suggest that they do (Oakhill, 1982; Oakhill et al., 1986) may have arisen from the heavy memory demands of the procedures

used. However, when using the version of the sentence recognition test that incorporated heavy memory demands, identical to the procedure used by Oakhill and colleagues (Oakhill, 1982; Oakhill et al., 1986), Experiments 2 to 5 continued to reveal no group differences. The work presented here provides no evidence that less-skilled comprehenders are less likely than skilled comprehenders to integrate information, even under heavy memory demands.

The discrepancy between this finding and those by Oakhill and colleagues (Oakhill, 1982; Oakhill et al., 1986) has two possible explanations. One possibility is that the children who participated in the experiments by Oakhill and colleagues were actually identifying semantically congruent foils as familiar on the basis of semantic congruency. It is possible that seven- and eight-year-olds might consider that a description of something that was heard before constitutes a sentence that was heard before. To avoid this possible misunderstanding, the children who participated in the work presented here were explicitly instructed not to identify such 'trick' sentences as old, and were administered practice trials to verify that they understood the instructions. The second possible explanation for the different findings is a success of recent education strategies. An emphasis on literacy education has been recently instated in British schools, and it has been compulsory to spend one hour each day focusing on a curriculum of literacy skills. Children may be more aware of the need to integrate story information than they were twenty years ago. Whilst some children still exhibit low comprehension performance with large pieces of text, such as in the Neale Analysis of Reading Ability, they may now be able to implement these new skills with short simple materials, such as those used in the sentence recognition test, and be more able to do so than the less-skilled comprehenders identified by Oakhill and colleagues nearly 20 years ago.

The second aim of Experiments 1 to 5 was to investigate whether poor memory ability might underlie low comprehension ability. The evidence discussed so far suggests that less-skilled comprehenders were not only just as likely as skilled comprehenders to integrate information, but were also as able to maintain the integrated information, providing no evidence that skilled and less-skilled

comprehenders differ in memory for story information. In addition, throughout Experiments 2 to 5, the comprehension groups were no differently affected in their overall accuracy by the increased memory demands between the two presentation conditions. Furthermore, Experiment 5 provided evidence that the severely impaired comprehension group did not exhibit impaired performance on two complex working memory tests, although children's span was low, and might be considered at floor. Experiments 2 to 5 provided no evidence that less-skilled comprehenders exhibited weaker memory ability than skilled comprehenders. This is contrary to existing evidence suggesting that comprehension ability is strongly related to performance on complex working memory tasks, i.e., tasks that require concurrent processing and storage of information, and that child comprehension is related to simple storage ability (see Sections 1.2.3 and 1.4.3).

The current findings do not agree with some existing findings. One conclusion from this evidence might be that the Neale Analysis of Reading Ability (Neale, 1997), by which the comprehension groups were selected, did not provide a reliable assessment of comprehension ability. One major concern about the Neale Analysis of Reading Ability is that comprehension ability is assessed by open-ended questions that require children to describe what they have understood. Such questions require – in addition to comprehension – meta-comprehension, sufficient expressive speech, and a certain amount of confidence. It is therefore possible that by using the Neale Analysis of Reading Ability to select reading groups in the current work, the actual comprehension ability of the less-skilled comprehension groups was underestimated, accounting for the lack of evidence that the groups differed on cognitive skills supposed to underlie comprehension ability.

In light of concerns about the Neale Analysis of Reading Ability, the findings can be restated – Experiments 1 to 5 found that children who differed in their ability to answer questions about what they have read did not differ in their tendency to integrate information in what they heard. There were differences between the demands of the two tasks – the comprehension assessment in the

Neale Analysis of Reading Ability and the sentence recognition test – that might account for this pattern of results. The first is a difference in presentation condition. This assumes that poor reading comprehension extends to poor listening comprehension. However, a strong body of evidence suggests that this is indeed the case (e.g., Stothard & Hulme, 1992; Carr et al., 1990; Nation & Snowling 1997)

Another possible explanation of why groups differed on one measure but not the other, is that the low performance of poor comprehenders might be limited to lengthy and complex language structures. In the present experiments, children were selected as poor comprehenders on the basis of their performance on lengthy and complex written prose in the Neale Analysis of Reading Ability (Neale, 1997) but their performance on the cognitive skills supposed to underlie comprehension ability was assessed on short, simple text.

In conclusion, across a series of experiments, it was found that children who were matched for reading accuracy but who differed in their ability to answer questions about what they read, did not exhibit difference in their tendency to integrate or maintain information from aurally presented short stories. The work presented in this chapter, therefore, provides no support for the hypotheses that reduced integration of information or poor memory ability underlie poor comprehension.

CHAPTER 3

METHODOLOGICAL MANIPULATIONS OF COMPREHENSION

In the five experiments presented in Chapter 2 groups of skilled and less-skilled comprehenders were selected on the basis of their performance on the reading accuracy and reading comprehension measures of the Neale Analysis of Reading Ability (Neale, 1997). Their performance on a sentence recognition task was then compared, to examine how the groups differed in their tendency to integrate information, and their ability to maintain that integrated information. In Chapter 2, no evidence was found that the groups differed on either count.

This finding was surprising, since Oakhill and colleagues (Oakhill, 1982; Oakhill et al., 1986) did find a group difference on an identical task. Some possible explanations for this discrepancy in results were provided in Chapter 2. The results were even more surprising because of the lack of evidence that skilled and less-skilled comprehenders were differentially affected by memory load, both in the different memory conditions of the sentence recognition test in Experiments 2 to 5, and explicitly in two complex working memory tasks in Experiment 5. These findings are contrary to existing evidence that comprehension ability relates to performance on complex working memory tasks (see Section 1.2.3 and 1.4.3).

The results of Chapter 2 could be accounted for by inadequate group selection, which in turn gives rise to concerns about the validity of the comprehension assessment provided by the Neale Analysis of Reading Ability. One concern was that the Neale Analysis of Reading Ability confounds reading accuracy and comprehension, and this point is considered in great detail in Chapter 4. Another concern lies in the type of question that is used to assess comprehension, and specifically that the questions make demands that are additional to comprehension, and might thereby underestimate the comprehension of some children.

In the Neale Analysis of Reading Ability, children are assessed on their ability to provide descriptive answers to open-ended questions about their comprehension – referred to in this work as *tell me* questions. In addition to sufficient comprehension, such questions require explicit awareness of how the text information fitted together and inferred other information, and ability to construct an answer that describes and explains that information. These extra demands run the risk that children who understood the text well might still answer the questions poorly because of poor meta-comprehension, or poor expressive speech, or even low confidence. In contrast, it might be argued that *true/false* questions provide a more pure measure of comprehension: Children compare their representation of the recognition sentence to their representation of the story information, then make a forced choice response, depending on whether these two representations match.

Experimental evidence exists to support the suggestion that the deficit in some children identified by the Neale Analysis of Reading Ability as less-skilled comprehenders might lie in formulating answers to questions about their comprehension. In two studies that examined ability to fill in missing information – either marked by a space or a non-word – the results depended on the response required. Cain, Oakhill, and Sharp (1998) found that less-skilled comprehenders were less able than skilled comprehenders to derive the meaning of a non-word, when children were required to explain the meaning. But Nation and Snowling (1997) found that ability to fill in a missing word was not dependent on comprehension ability, when the response that children had to make was selecting the appropriate word from a choice of four.

In Experiment 6, it was investigated whether the traditional *tell me* questions of the Neale Analysis of Reading Ability can underestimate comprehension. The performance of skilled and less-skilled comprehenders – as identified by the Neale Analysis of Reading Ability – was compared on a parallel version of the test where questions had been transformed to forced-choice *true/false* decisions.

Chapter 2 also raised the question of whether comprehension groups as selected by the Neale Analysis of Reading Ability did not differ on the sentence recognition test because these two tests placed different demands on children. It was suggested that the sentence recognition test may have been insufficiently sensitive to cognitive differences between children who differed on the comprehension measure of the Neale Analysis of Reading Ability. Firstly, the two tests differed in the type of response required – descriptive answers in the Neale Analysis of Reading Ability, and forced choice in the sentence recognition test. This difference was discussed above. Secondly, the majority of texts used in the Neale Analysis of Reading Ability were lengthy and complex written prose, while the stories used in the sentence recognition test were short and simple. This difference points to the possibility that the Neale Analysis of Reading Ability identified children whose comprehension difficulties were limited to lengthy and complex language structures. If this were the case, such children would not be expected to exhibit cognitive differences from skilled comprehenders across the short and simple stories of the sentence recognition test. It was therefore investigated in Experiment 6 whether groups of skilled and less-skilled comprehenders – identified by the Neale Analysis of Reading Ability – differed on short simple texts as well as long complex texts.

Thirdly, the sentence recognition test and the Neale Analysis of Reading Ability differed in their condition of presentation, opening the possibility that the comprehension groups differed in one test but not the other because their comprehension differences were limited to oral reading. Chapter 3, therefore, also presents work investigating the generality of comprehension ability and performance across presentation condition. Firstly, it was investigated in Experiment 6 whether general comprehension ability extended specifically to silent reading comprehension. The evidence of strong generality of comprehension ability between listening and oral reading comprehension has been well documented (e.g., Daneman & Carpenter, 1980; Nation & Snowling, 1997; Stothard & Hulme, 1992). It has also been specifically demonstrated that individuals who perform poorly on oral reading comprehension also perform poorly on listening comprehension, in children (Carr et al., 1990) and adults

(Gernsbacher, Varner, & Faust, 1990). One might therefore expect general comprehension ability to extend to silent reading comprehension. However, oral reading (when supervised) and listening comprehension provide children with the words, while silent reading relies on reading accuracy. It was investigated in Experiment 6 whether oral reading and listening comprehension ability generalises to silent reading comprehension.

In light of concerns about the validity of the comprehension assessment provided by the Neale Analysis of Reading Ability, the relationship between performance on this test and a comprehension assessment that more closely resembled the real-life situations that children are in when reading for comprehension was also investigated in Experiment 6. This real-life comprehension measure assessed children in their usual classroom places, reading independently and at their own pace. This avoided some of the potentially intimidating factors inherent in administration of the Neale Analysis of Reading Ability. Furthermore, comprehension was assessed by *true/false* questions, thereby also avoiding some of the additional demands that the *tell me* questions make in the Neale Analysis of Reading Ability. In Experiment 6, three groups of children were selected with specific profiles across the Neale Analysis of Reading Ability, and their performances on the real-life comprehension assessment were compared. These groups are those of skilled readers (referred to in Chapter 2 as skilled comprehenders) and less-skilled comprehenders, and less-skilled decoders, with low decoding (reading accuracy) and good comprehension score, to provide an indication of how a specific deficit in reading accuracy affected performance on tasks.

Furthermore, it was explored in Experiments 6 and 7 whether the comprehension of the three different reading groups might be differentially affected by the change in condition from reading aloud in the Neale Analysis of Reading Ability to reading silently in the classroom test. Direct comparisons were conducted between oral reading comprehension on the *true/false* version of the Neale Analysis of Reading Ability described above, and unaided silent reading comprehension, with *true/false* questions. Whilst both oral and silent

reading require identification of the meaning of words, oral reading also requires retrieval and vocalisation of the sounds of words. Children who find reading aloud for meaning difficult – because they find either decoding or comprehension difficult – might be aided by the removal of the vocalisation demands of the task. On the other hand, reading aloud might force less-skilled decoders to read each word, rather than accidentally skipping some or deliberately avoiding problematic ones. It might also make errors apparent, as the reader hears what has been decoded. In addition, in some oral reading tasks, such as the Neale Analysis of Reading Ability (Neale, 1997), errors are corrected, therefore providing children with the correct words. These propositions would predict that reading aloud would improve text comprehension, by providing the reader with more accurate information.

In summary, the two experiments reported in this chapter investigated whether comprehension groups identified in a typical manner by performance on the Neale Analysis of Reading Ability continued to differ across manipulations of question type, length and complexity of text, and presentation condition. The performance of decoding groups was also investigated across presentation condition.

3.1 Experiment 6

3.1.1 Method

Participants

Two-hundred and twelve Year 3 children from six urban Bristol schools participated in Experiment 6. The sample ranged in age from 7.17 years to 8.58 years, with mean age 7.98 years, and standard deviation 0.32 years. The whole sample was screened on the Neale Analysis of Reading Ability, form 1, (Neale, 1997). Sixteen skilled readers were selected according to the following criteria (a) at least age-appropriate reading accuracy and comprehension, (b) reading accuracy and comprehension ages no more than two years above actual age, and (c) reading accuracy and comprehension ages within 12 months of each

other. Sixteen less-skilled comprehenders were selected according to the following reading profile: (a) at least age-appropriate reading accuracy, and (b) reading comprehension age at least 12 months below actual age. Sixteen less-skilled decoders were selected to the criteria (a) good reading comprehension, and (b) reading accuracy age at least 6 months below age. Unfortunately, comprehension ability cannot be reliably assessed by the Neale Analysis of Reading Ability in children with particularly poor reading ability – in order to have the opportunity to obtain at least age-appropriate comprehension age, Year 3 children must progress to Level 3 of the Neale. But, poor readers frequently reach the cut-off number of reading errors before Level 3, and are therefore not permitted to attempt Level 3, including its comprehension questions. To permit identification of less-skilled decoders despite the unfair restrictions on their comprehension performance, any child whose reading was not good enough to progress onto Level 3, but who had made a maximum of two comprehension errors in total across Levels 1 and 2 was included. Children whose reading was not good enough to even progress onto Level 2 were not included, as their reading ability was considered to be at floor, and too poor to participate in the silent reading passages. Table 3.1 summarises the profiles of the three reading groups and the whole sample.

Table 3.1. Participant profiles in Experiment 6. Ages in years. Standard deviations in parentheses.

			Neale Analysis of Reading Ability reading ages	
	N	Age	Accuracy	Comprehension
Whole sample	212	8.0 (0.32)	8.5 (1.59)	7.9 (1.36)
Skilled readers	16	7.9 (0.23)	8.4 (0.32)	8.5 (0.43)
Less-skilled comprehenders	16	7.9 (0.34)	9.0 (1.03)	6.5 (0.41)
Less-skilled decoders	16	8.1 (0.31)	6.8 (0.56)	7.6 (0.72)

Materials

Four tests were administered. The Neale Analysis of Reading Ability, form 1 (Neale, 1997) yielded standardised reading accuracy and reading

comprehension scores for group selection, and the reading comprehension raw scores provided the measure of *aloud tell me* comprehension. The Neale Analysis of Reading Ability, form 2, (Neale, 1997) formed the basis for the second test. While the stories were used in their original format, the questions were changed into *true/false* statements. The content of original questions and their correct answers was closely adhered to. In a few cases this was not possible, and new *true/false* statements were constructed to tap the same story information. Half of the *true/false* statements following each story tapped information that had been stated explicitly within the text, and half tapped information that needed to be inferred from text information. This version of the Neale Analysis of Reading Ability provided the measure of *aloud true/false* performance. The listening and unaided reading comprehension tests, developed for this thesis (see Section 4.1), were also administered, providing the measures of *listening true/false* comprehension, and *silent true/false* comprehension respectively.

Scoring

Performance on all comprehension tests was scored as the number of questions correct. The four tests each had four questions at Level 1, and eight questions per story at Levels 2 to 6. Where analyses required between-test comparisons, there was a concern that differences in performance might have arisen from a different number of questions being attempted on each test, due to accuracy cut-off criteria in administration of the *aloud tell me* and *aloud true/false* versions of the Neale Analysis of Reading Ability. To ensure that performance was being compared across the same number of stories for each child, a 'matched level' was derived for each child. This was the maximum level at which the child completed all of the *aloud tell me*, *aloud true/false* and the *silent true/false* tests. Scores were calculated to this level, referred to as *matched scores*, matched within subjects across tests. The term *absolute score* refers to the score obtained on any test, calculated over all the questions attempted in that test.

Procedure

Children were administered the unaided reading comprehension test in their usual classes, reading passages of increasing difficulty silently, independently, and at their own pace, turning the page after each story to read and tick or cross *true/false* statements about that story. Questions were answered from memory. Children were also administered the listening comprehension test in their usual classes, listening to passages of increasing difficulty, and ticking or crossing *true/false* statements that they heard about the story. These questions were also answered from memory. (See Section 4.1.4 for detailed description of the listening and unaided comprehension tests.) Children were then administered the Neale Analysis of Reading Ability, form 1, in a quiet area of the school. Children read passages of increasing difficulty aloud to the experimenter. Errors were corrected and counted. If no more than sixteen errors were made, questions were asked about the story, and the number of correct answers noted. On the basis of performance on the Neale Analysis of Reading Ability, form 1, groups were selected as described above. These children were then administered the transformed version of the Neale Analysis of Reading Ability, form 2, which was also read aloud by children and corrected by the experimenter. If no more than sixteen errors were made, *true/false* questions were asked about the story. Responses were noted at testing.

3.1.2 Results

Skilled readers and less-skilled comprehenders completed more stories than less-skilled decoders. Skilled readers scored better than both other groups on all tests. Less-skilled comprehenders scored below both other groups in the *aloud tell me* condition and less-skilled decoders scored below both other groups in the *aloud true/false* and *silent true/false* conditions. Table 3.2 summarises group performances (the maximum possible score is 44 on each test).

Table 3.2. Mean comprehension scores in Experiment 6, with standard deviations in parentheses.

		Matched scores		
	Matched level	Aloud tell me	Aloud true/false	Silent true/false
Skilled readers	3.8 (0.40)	17.7 (2.30)	17.8 (3.69)	22.6 (2.92)
Less-skilled comprehenders	4.2 (0.83)	7.3 (2.27)	16.8 (2.56)	21.9 (4.19)
Less-skilled decoders	2.4 (0.63)	13.4 (3.67)	11.4 (3.84)	11.4 (5.75)

The first stage of analysis investigated whether groups of children who differed on the comprehension measure of the Neale Analysis of Reading Ability continued to differ when the questions were replaced with *true/false* questions. In order to validate comparisons between the performances of the skilled readers and less-skilled comprehenders, it was first determined whether the groups were assessed across the same number of questions. A one-way analysis of variance, with group as factor, was conducted on matched levels obtained by skilled readers and less-skilled comprehenders (see the section on scoring for information). The analysis revealed non-significant main effect of group, $F(1,30)=2.62$, $p>.05$.

Having ascertained that skilled readers and less-skilled comprehenders had attempted the same number of questions, the next set of analyses were conducted to ascertain whether comprehension groups were differentially affected by question type. One-way analyses of variance, with group as factor, were conducted on matched comprehension scores (see the section on scoring for information) obtained by skilled readers and less-skilled comprehenders in the *aloud tell me* and *aloud true/false* tests separately. The analyses revealed significant main effect of group in the *aloud tell me* condition, with more correct answers provided by skilled readers than by less-skilled comprehenders, $F(1,30)=167.13$, $p<.001$, and non-significant main effect of group in the *aloud true/false* test, $F(1,30)<1$.

To examine how the comprehension groups differed as the stories became longer and more complex, the comprehension performance of skilled readers and less-skilled comprehenders was compared at each level of difficulty, with both question types. All of the skilled readers and less-skilled comprehenders completed Level 3 on both *aloud tell me* and *aloud true/false* tests, but a small number of each group did not progress onto Level 4. These children were omitted from the level-by-level analyses, leaving 13 children in each comprehension group who completed all stories at Levels 1,2,3 and 4. Table 3.3 summarises the performance of these children at each level with each question type. Scores have been transformed to percent correct for comparison.

Table 3.3. Summary of performance by children who completed Levels 1 to 4 on *aloud tell me* and *aloud true/false* in Experiment 6. Percentage correct, with standard deviations in parentheses.

Question type	Group	Level 1	Level 2	Level 3	Level 4
Open-ended	Skilled readers	90.4 (16.3)	91.4 (12.9)	61.5 (17.3)	29.8 (12.0)
	Less-skilled comprehenders	57.7 (23.7)	46.2 (20.7)	12.5 (12.5)	5.8 (8.3)
Forced-choice	Skilled readers	76.9 (16.0)	77.9 (12.7)	55.8 (23.2)	60.6 (22.2)
	Less-skilled comprehenders	75.0 (17.7)	72.1 (17.8)	37.5 (13.5)	43.3 (13.1)

A three-way analysis of variance, with group, question type and level as factors was conducted on scores (percentage correct) attained by these skilled readers and less-skilled comprehenders in the *aloud tell me* and *aloud true/false* conditions. The analysis revealed significant main effect of question type, $F(1,24)=37.53$, $p<.001$, with more correct answers to forced-choice questions than open questions, and significant main effect of group, $F(1,24)=72.54$, $p<.001$, with more correct answers provided by skilled readers than less-skilled comprehenders. The analysis also revealed significant main effect of level, $F(3,72)=78.34$, $p<.001$. Post-hoc Tukey analysis on scores across levels revealed that while similar scores were attained at levels 1 and 2 and at levels 3 and 4, higher scores were attained at levels 1 and 2 than at levels 3 and 4. The analysis of variance also revealed significant interactions between question type

and level, F(3,72)=10.66, and between question type and group, F(1,24)=40.36, both p<.001. Post-hoc Tukey analysis on scores across question type and levels revealed that higher scores were attained at levels 1 and 2 than at levels 3 and 4 with both question types, and that similar scores were attained across question type at all levels except level 4, where forced-choice questions yielded higher scores than open-ended questions. Post-hoc Tukey analysis on scores across question type and group revealed that while question type did not affect the performance of the skilled readers, less-skilled comprehenders provided more correct answers to forced-choice questions than to open-ended questions. The post-hoc analysis also revealed that skilled readers attained higher scores than less-skilled comprehenders in the open-ended condition but not the forced-choice condition. The analysis of variance revealed non-significant interactions between level and group, F(3,72)=2.34, and between level, group and question type, F(3,72)=2.34, both p>.05.

The next stage of analysis investigated whether general comprehension ability extended to silent reading comprehension. Three measures of comprehension were obtained for the whole sample of children who participated in Experiment 6 – *aloud tell me* comprehension, *listening true/false* comprehension, and *silent true/false* comprehension. A series of Pearson correlations were conducted, partialling out both age and accuracy score obtained on the Neale Analysis of Reading Ability, and revealed significant correlations between comprehension in the *silent true/false* and *aloud tell me* conditions, r= .32, p<.001, between *silent true/false* and *listening true/false* conditions, r= .38, p<.001, and between *aloud tell me* and *listening true/false* conditions, r= .52, p<.001, all N=212, beyond age and reading accuracy skill.

The next stage of analysis investigated how the assessment of reading and comprehension ability provided by the Neale Analysis of Reading Ability related to independent reading comprehension. The first analyses compared performance of the comprehension groups as selected by the Neale Analysis of Reading Ability. To verify the selection of these groups, a one-way analysis of variance, with group as factor, was conducted on absolute comprehension

scores (see the section on scoring for information) obtained by skilled readers and less-skilled comprehenders on the Neale Analysis of Reading Ability. This analysis revealed significant main effect of group, with more correct answers provided by skilled readers than by less-skilled comprehenders, $F(1,30)=193.14$, $p<.001$. On the unaided reading comprehension measure (maximum score 44), skilled readers obtained mean absolute score of 31.94, (standard deviation 3.07) and less-skilled comprehenders obtained mean absolute score of 29.94, (standard deviation 3.91). A one-way analysis of variance, with group as factor, was conducted on these absolute comprehension scores obtained by skilled readers and less-skilled comprehenders on the unaided reading comprehension test. The analysis revealed non-significant main effect of group, $F(1,30)=2.25$, $p>.05$.

The next analyses compared performance of the decoding groups as selected by the Neale Analysis of Reading Ability. To verify the selection of these groups, a one-way analysis of variance, with group as factor, was conducted on absolute comprehension scores obtained by skilled readers and less-skilled decoders on the Neale Analysis of Reading Ability. This analysis revealed significant main effect of group, with more correct answers provided by skilled readers than by less-skilled decoders, $F(1,30)=193.14$, $p<.001$. To verify that this difference resulted from the assessment of skilled readers over more questions than less-skilled decoders, comprehension score obtained by these two groups on the Neale Analysis of Reading Ability was calculated over Levels 1 and 2, which all skilled readers and less-skilled decoders completed. Over Levels 1 and 2 of the Neale Analysis of Reading Ability (maximum score 12), skilled readers obtained mean comprehension score 11.00 (standard deviation 0.30) and less-skilled decoders obtained mean score 11.13 (standard deviation 0.20). To verify that the decoding groups did not differ in comprehension performance when they were assessed over the same number of questions, a one-way analysis of variance, with group as factor, was conducted on comprehension scores obtained by skilled readers and less-skilled decoders on the standard form of the Neale Analysis of Reading Ability, with traditional

tell me questions. The analysis revealed non-significant main effect of group, $F(1,30)<1$.

On the unaided reading comprehension test (maximum score 44), less-skilled decoders obtained mean absolute score of 25.00, (standard deviation 4.63). A one-way analysis of variance, with group as factor, was conducted on absolute comprehension scores obtained by skilled readers and less-skilled decoders on the unaided reading comprehension test. The analysis revealed significant main effect of group, with more correct answers provided by skilled readers than by less-skilled decoders, $F(1,30)=24.95$, $p<.001$.

The final stage of analysis investigated whether the comprehension of children with less-skilled decoding or less-skilled comprehension ability on the Neale Analysis of Reading Ability was aided or impaired by removing the extra vocalising demands of reading aloud. A series of paired sample t-tests was conducted on matched scores obtained in the *aloud true/false* and *silent true/false* conditions, by each group separately. The analyses revealed that skilled readers and less-skilled comprehenders answered significantly more correct questions when reading silently than when reading aloud, $t(15)=8.39$ and $t(15)=7.20$ respectively, both $p<.001$, and that the number of correct answers provided by less-skilled decoders did not differ between condition, $t(15)=0.00$.

To examine whether groups who differed in comprehension ability were differentially affected by reading condition, a two-factor analysis of variance, with condition and group as factors, was conducted on matched scores obtained in the *aloud true/false* and *silent true/false* conditions, by skilled readers and less-skilled comprehenders. The analysis revealed significant main effect of condition, with more correct answers provided when reading silently than when reading aloud, $F(1,30)=117.76$, $p<.001$, non-significant main effect of group, $F(1,30)<1$, and non-significant interaction between condition and group, $F(1,30)<1$.

To validate direct comparison between the performance of skilled readers and less-skilled decoders across the two conditions, it was first investigated whether the matched scores of the groups were calculated over the same number of questions. A one-way analysis of variance, with group as factor, was conducted on the matched levels obtained by skilled readers and less-skilled decoders. The analysis revealed significant main effect of group, with more levels completed by skilled readers than by less-skilled decoders, $F(1,30)=54.18$, $p<.001$. To enable direct comparison between the groups, scores were calculated over Levels 1 and 2, which all children completed. To Level 2 on *aloud true/false* (maximum score 12), skilled readers obtained mean score 9.56 (standard deviation 1.21), and less-skilled decoders obtained mean score 9.25 (standard deviation 1.69). To Level 2 on *silent true/false* (maximum score 12), skilled readers obtained mean score 10.44 (standard deviation 1.31), and less-skilled decoders obtained mean score 8.31 (standard deviation 2.15).

To examine whether the decoding groups were differentially affected by reading condition, the two-way analysis of variance, with condition and group as factors, was conducted on scores calculated to Level 2 for skilled readers and less-skilled decoders in the *aloud true/false* and *silent true/false* conditions. The analysis revealed non-significant main effect of condition, $F(1,30)<1$, and significant main effect of group, $F(1,30)=8.49$, $p<.01$, with more correct answers provided by skilled readers than by less-skilled decoders, and significant interaction between group and condition, $F(1,30)=5.16$, $p<.05$. To explore the significant interaction between group and condition, a series of post-hoc Tukey comparisons was conducted on scores obtained to Level 2 by each group in each condition. The comparisons revealed that skilled readers provided significantly more correct answers than less-skilled decoders when reading silently, $p<.01$, but that the groups did not differ when reading aloud.

In summary, when reading aloud, skilled readers provided more correct answers than less-skilled comprehenders to *tell me* questions, but not to *true/false* questions, over the range of stories. Although skilled readers were more accurate than less-skilled comprehenders at every level of difficulty when

assessed by *tell me* questions, the groups did not differ on the easier stories when assessed by *true/false* questions. Experiment 6 also demonstrated that silent reading comprehension correlated highly with both oral reading comprehension and listening comprehension, beyond age and accuracy, across a large population sample. Compared to the performance of skilled readers, the Neale Analysis of Reading Ability underestimated the everyday reading comprehension of less-skilled comprehenders, but overestimated the everyday reading comprehension of less-skilled decoders. In addition, skilled readers and less-skilled comprehenders exhibited better comprehension when reading silently than when reading aloud, but less-skilled decoders were not aided by the change in condition.

3.2 Experiment 7

In Experiment 6, comprehension performance was compared between oral and silent reading conditions, in groups of different reading profiles, using *true/false* questions. In Experiment 7, this investigation was repeated, using *tell me* questions, to explore the effect of question type and reading condition on reading comprehension.

3.2.1 Method

Participants

The children who participated in Experiment 1 also participated in Experiment 7. In addition to the sixteen skilled readers (referred to in Experiment 1 as skilled comprehenders) and the sixteen less-skilled comprehenders, sixteen less-skilled decoders were selected according to the criteria on the Neale Analysis of Reading Ability: (a) comprehension age above chronological age, (b) accuracy age at least six months below reading comprehension age. The less-skilled decoding group had mean chronological age 8.08 years (standard deviation 0.22 years), mean reading accuracy age 8.01 years (standard deviation 0.55 years), and mean reading comprehension age 9.29 years (standard deviation 0.70 years).

Design

Two factor mixed design, with three levels of between-subject factor reading group (skilled readers, less-skilled comprehenders, less-skilled decoders) and two levels of within-subject factor condition (aloud, silent). Dependent variable was reading comprehension score, as a direct count of comprehension questions answered correctly.

Materials

Two tests were administered. The Neale Analysis of Reading Ability, form 2 (Neale et al., 1989) yielded standardised reading accuracy and reading comprehension scores for group selection, and the reading comprehension raw scores provided the measure of *aloud* comprehension. The Neale Analysis of Reading Ability, form 1, (Neale et al., 1989) was also administered, and provided the measure of *silent* comprehension.

Procedure

Children were administered the Neale Analysis of Reading Ability, form 2, in a quiet area of the school. Children read passages of increasing difficulty aloud to the experimenter. Errors were corrected and counted. If no more than sixteen errors were made, questions were asked about the story, and responses noted at testing. On the basis of performance on the Neale Analysis of Reading Ability, form 2, groups were selected as described above. These children were then administered the Neale Analysis of Reading Ability, form 1, which was read silently. After each story, the experimenter asked the comprehension questions. Responses were noted at testing. Children completed the same number of stories on form 1 of the Neale Analysis of Reading Ability as were completed on form 2 during screening.

3.2.2 Results

Less-skilled comprehenders obtained lower scores than both skilled readers and less-skilled decoders in both conditions. Less-skilled decoders obtained lower scores when reading silently than when reading aloud. Table 3.4 summarises

group performance in the two conditions. Maximum level is six and maximum possible score in each condition is forty-four.

The first analyses were conducted within each group, to investigate whether any group was affected by reading condition. A series of paired-sample t-test comparisons was conducted between the scores obtained in each condition, by each group separately. The analyses revealed that skilled readers and less-skilled comprehenders were not affected by reading condition, $t(15)=0.27$, $t(15)=0.00$ respectively, both $p>.05$, and that less-skilled decoders answered significantly more questions correct after reading aloud than after reading silently, $t(15)=4.52$, $p<.001$.

Table 3.4. Mean level obtained by each group, and mean comprehension scores obtained in each reading condition in Experiment 7. Standard deviations in parentheses.

	Max Level	Aloud	Silent
Skilled readers	4.5 (0.73)	18.9 (3.09)	19.4 (8.85)
Less-skilled comprehenders	4.7 (0.79)	12.8 (1.91)	12.8 (3.84)
Less-skilled decoders	3.9 (0.44)	19.6 (2.50)	16.1 (3.07)

The next set of analyses explored the performance of the comprehension groups. In order to validate direct comparison between the group scores, it was first ascertained whether the groups were assessed over the same number of questions. The one-way analysis of variance, with group as factor, was conducted on the maximum levels obtained by skilled readers and less-skilled comprehenders. The analysis revealed non-significant main effect of group, $F(1,30)<1$. The next analysis investigated whether less-skilled comprehenders continued to obtain lower scores than skilled readers when reading silently. A one-way analysis, with group as factor, was conducted on scores obtained by skilled readers and less-skilled comprehenders after reading silently. The analysis revealed significant main effect, $F(1,30)=7.55$, $p<.05$, with more correct answers provided by skilled readers than by less-skilled comprehenders.

The next analysis investigated whether the comprehension groups were differentially affected by reading condition. The two-way analysis of variance, with group and reading condition as factors, was conducted on the comprehension scores obtained by skilled readers and less-skilled comprehenders. The analysis revealed non-significant main effect of condition, $F(1,30)<1$, non-significant interaction between condition and group, $F(1,30)<1$, and significant main effect of group, $F(1,30)=16.77$, $p<.001$, with more correct answers provided by skilled readers than by less-skilled comprehenders.

The final stage of analysis explored the performance of the decoding groups. The first analysis investigated whether the skilled readers and less-skilled decoders were assessed over the same number of questions. A one-way analysis of variance, with group as factor, was conducted on the maximum level obtained by skilled readers and less-skilled decoders. The analysis revealed significant main effect of group, with skilled readers completing more levels than less-skilled decoders, $F(1,30)=6.94$, $p<.05$. In order to validate direct comparison between groups, scores were calculated to Level 3 (maximum score 20) which all skilled readers and less-skilled decoders completed. Skilled readers obtained mean score 14.19 (standard deviation 1.72) when reading aloud, and mean score 14.31 (standard deviation 4.03) when reading silently. Less-skilled decoders obtained mean score 15.19 (standard deviation 1.83) when reading aloud, and mean score 13.88 (standard deviation 1.93) when reading silently.

The next analysis investigated whether less-skilled decoders continued to obtain similar comprehension scores to skilled readers when reading silently. A one-way analysis of variance, with group as factor, was conducted on scores obtained by skilled readers and less-skilled decoders after reading silently, to Level 3 only. The analysis revealed non-significant main effect of group, $F(1,30)<1$. The final analysis investigated whether the decoding groups were differentially affected by reading condition. A two-way analysis of variance, with group and condition as factors, was conducted on scores obtained by skilled readers and less-skilled decoders, to Level 3 only. The analysis revealed

non-significant main effect of condition, $F(1,30)=1.33$, $p>.05$, non-significant interaction between condition and group, $F(1,30)=1.94$, $p>.05$, and non-significant main effect of group, $F(1,30)<1$.

In summary, skilled readers and less-skilled comprehenders were unaffected by reading condition, but less-skilled decoders obtained lower scores when reading silently than when reading aloud.

3.3 General Discussion

Skilled readers and less-skilled comprehenders, matched on decoding ability but differing on comprehension ability, and a group of less-skilled decoders, with good comprehension ability, were selected on the basis of performance on the Neale Analysis of Reading Ability (Neale, 1997; Neale et al., 1989). These groups were then compared across manipulations of question type, length and complexity of materials, and presentation condition.

Although the groups of skilled readers and less-skilled comprehenders were selected by their good and poor performance respectively on the comprehension assessment of the Neale Analysis of Reading Ability, in its traditional format with *tell me* questions, it was found in Experiment 6 that these groups no longer differed in comprehension performance when assessed by ability to answer *true/false* questions. Furthermore, whilst it was found in Experiment 6 that the less-skilled comprehension group scored well on a silent reading test with *true/false* questions, it was found in Experiment 7 that a group of less-skilled comprehenders scored poorly on an equivalent silent reading test that featured *tell me* questions. These findings suggest that groups of children selected as less-skilled comprehenders on the basis of low performance on the Neale Analysis of Reading Ability might be polluted by children who have a difficulty in the extra demands made by the *tell me* questions – meta-comprehension, expressive speech, confidence – rather than comprehension per se. In short, the findings demonstrate that the traditional open-ended questions used in the Neale Analysis of Reading Ability might underestimate children's internal comprehension of text.

The effect of length and complexity of text material on comprehension performance was also investigated in Experiment 6. Scores obtained by skilled and less-skilled comprehenders – as selected by performance on the traditional version of the Neale Analysis of Reading Ability – were compared across the levels of difficulty of the traditional and alternative versions of the Neale Analysis of Reading Ability. The groups differed at every level of difficulty on the traditional version of the test. When the question demands were reduced, with comprehension assessed by ability to answer *true/false* questions, the comprehension groups continued to differ on the more difficult stories, but they not differ on the shorter, more simple levels. This finding implies that the less-skilled comprehenders – as selected by the Neale Analysis of Reading Ability – were not impaired on their comprehension of short, simple texts.

The finding that children who differ on the comprehension measure of the Neale Analysis of Reading Ability did not differ in their comprehension of short, simple stories has important implications for the results of Experiments 1 to 5. These experiments used short, simple stories to investigate the extent to which skilled and less-skilled comprehenders – as identified by the Neale Analysis of Reading Ability –differed in their tendency to integrate information, and did not find evidence that they did. The findings of Experiment 6 suggest that the materials used were too short and simple to detect cognitive differences between the comprehension groups. The important implications for future research would be that, in order to provide a reliable measure of cognitive abilities underlying comprehension, materials should be constructed that are no easier than the Level 3 stories of the Neale Analysis of Reading Ability.

The generality of comprehension across modality was also investigated in Experiment 6. Existing work has demonstrated that individual comprehension ability transcends condition, so that individuals with good oral reading comprehension also exhibit good listening comprehension (Carr et al., 1990; Nation & Snowling, 1997; Stothard & Hulme, 1992), and Gernsbacher et al. (1990) demonstrated consistency of comprehension between stories presented

in written, aural and picture format. Further evidence was provided in Experiment 6 that oral reading comprehension and listening comprehension correspond highly within the general population. Furthermore, it was demonstrated in Experiment 6 that independent silent reading comprehension corresponded highly with both oral reading comprehension and listening comprehension, within the general population. This finding remained robust beyond contributions of age and reading accuracy. These findings support the suggestion made by Gernsbacher et al. (1990) that common cognitive processes underlie comprehension, and that comprehension strategies are not unique to certain modalities.

The relationship between the assessments provided by the Neale Analysis of Reading Ability and comprehension of texts read independently in a classroom setting was also investigated in Experiment 6. Skilled readers – children who performed well on both the reading accuracy and reading comprehension measures of the Neale Analysis of Reading Ability – were assumed to provide an indication of skilled reading performance on the classroom test. Less-skilled comprehenders, who performed poorly on the comprehension assessment and well on the reading accuracy assessment of the Neale Analysis of Reading Ability, did not comprehend poorly compared to skilled readers when reading silently and independently. In contrast, less-skilled decoders, who performed well on the comprehension assessment and poorly on the reading accuracy assessment or the Neale Analysis of Reading Ability, did comprehend poorly compared to skilled readers when reading silently and independently, presumably due to uncorrected reading errors. These findings reveal that the assessments provided by the Neale Analysis of Reading Ability might not provide a reliable indication of the reading comprehension of individuals in a classroom setting

The effect of the extra demands that reading aloud has on text comprehension, compared to reading silently, was also investigated in Experiments 6 and 7. The three reading groups of skilled readers, less-skilled comprehenders and less-skilled decoders, as selected on the basis of performance on the Neale Analysis

of Reading Ability, were compared between oral and silent reading comprehension tasks, by answering *true/false* questions from memory in Experiment 6, and by answering *tell me* questions from memory in Experiment 7. In Experiment 6, both less-skilled comprehenders and skilled readers provided more correct answers when reading silently than when reading aloud, whilst less-skilled decoders were not helped by reading condition. In Experiment 7, reading condition did not affect the skilled readers and less-skilled comprehenders, but less-skilled decoders performed more poorly when reading silently than when reading aloud.

The finding in Experiment 6 that skilled readers and less-skilled comprehenders answered more correct questions in the silent reading condition than in the oral reading condition has two possible explanations. The first is simply that the materials used in the silent condition were easier than the materials used in the oral reading condition. This possibility is supported by the finding that less-skilled decoders obtained lower scores when reading silently than when reading aloud in Experiment 7, but not in Experiment 6 – it appears that the handicap that silent reading poses to less-skilled decoders might have been balanced out by the easier materials in the silent condition of Experiment 6. This point could be examined by repeating the investigation with materials counterbalanced between conditions. The second possible explanation is that skilled readers and less-skilled comprehenders exhibited better comprehension in the silent reading test because of the change in condition. There are a number of ways in which the conditions of the two tests differ. The silent reading test was administered to children in their usual classes, at their usual places, surrounded by classmates and teacher, and this familiar setting might have optimised performance by assessing children when they felt safe and relaxed. This is in stark contrast to the potentially intimidating situation that children were in when assessed on the oral reading test, sitting alone and reading aloud under the watchful eye of an unknown experimenter. This possible explanation may also be supported by the findings in Experiment 7, where children were administered both the aloud and silent reading tests alone with the experimenter, and no group obtained higher scores when reading silently. This possibility could also be investigated, by

administering both conditions of the *true/false* tests used in Experiment 6 to individual children.

The second manner by which the conditions of the two tests differ is that the silent reading test gave children the opportunity to read at their own pace, and to re-read difficult or misinterpreted sections, whilst children tended to read the oral reading test straight through. The third difference between the test conditions is that the oral reading tests placed additional demands that the silent reading test did not, namely vocalising the text. The removal of this additional demand may have permitted greater concentration on comprehension, thereby improving comprehension.

Any one or any combination of these factors may have contributed to the improved comprehension exhibited by less-skilled comprehenders and skilled readers in Experiment 6, and one might have expected less-skilled decoders to have also performed better in the silent reading condition. However, the less-skilled comprehenders and skilled readers were selected to have good reading accuracy – they made few errors when reading aloud the passages in the Neale Analysis of Reading Ability. Less-skilled decoders, on the other hand, made many decoding errors. In the oral reading tests, such errors were corrected by the experimenter, providing less-skilled decoders with the text information from which to build comprehension, and less-skilled decoders exhibited good comprehension in this condition. However, in the silent reading condition, children worked independently. It follows that children with poor reading accuracy did not supply themselves with all the information presented to them by the text when they were unaided, impairing their comprehension. In short, less-skilled decoders answered fewer comprehension questions correctly than skilled readers in both Experiments 6 and 7 when reading silently and unaided, which might be accounted for by their low reading accuracy.

The practical implications arising from these findings are that children with good reading accuracy understand better if they do not have to vocalise what they read, and if they can work independently and at their own pace in a

comfortable setting. The comprehension of children with poor reading accuracy, on the other hand, suffers in independent work, and that such children need the help of another to fill in the gaps that their less-skilled decoding can otherwise leave. The implications for future research would be that experimental tests might be devised that implement silent reading in order to tap skills underlying comprehension ability, but only for use with children who exhibit good reading ability.

In summary, evidence was provided in Experiment 6 that the type of questions used in the Neale Analysis of Reading Ability might underestimate text comprehension, and that children who perform poorly on the comprehension measure of the Neale Analysis of Reading Ability do not necessarily comprehend easier materials poorly. Further support for the hypothesis that comprehension ability is general across modality was also provided in Experiment 6, but it was suggested that the Neale Analysis of Reading Ability should not be used to make inferences about the independent reading comprehension of children in a classroom setting. The patterns of performance across the combinations of reading condition and question type in Experiments 6 and 7 suggested that the comprehension performance of less-skilled comprehenders improves when assessed on *true/false* questions, and that the comprehension of less-skilled decoders suffers when required to read silently and unaided.

CHAPTER 4

READING AND COMPREHENSION AS SEPERABLE COMPONENTS OF THE NEALE ANALYSIS OF READING ABILITY

The Neale Analysis of Reading Ability (Neale, 1958; Neale et al., 1989; Neale, 1997) is a reading test that provides standardised scores of both reading accuracy and reading comprehension. Children read aloud stories, and the administrator counts and corrects any reading errors. When no more than 16 errors are made on a story, comprehension questions are asked. Children progress through stories of increasing difficultly until they make 16 or more reading errors on one story. The total number of reading errors made throughout the test yields a reading accuracy score, and the number of questions answered correctly yields a comprehension score.

The Neale Analysis of Reading Ability is a very useful tool of assessment in education, providing a measure of children's ability to explain their understanding of texts that they can read. This set of skills closely resembles successful reading for comprehension at school. This is particularly true of interactive situations, where children can ask for assistance, where questions can be rephrased when necessary, and where children's expressive speech is an important aspect of communication in the classroom. The Neale Analysis of Reading Ability, therefore, provides an important and valuable assessment of a child's successful reading for comprehension in an interactive educational situation. As a standardised and widely used test, it allows an individual child to be compared to the population, and, as the test has two parallel versions, can also be used to assess children's progress.

The Neale Analysis of Reading Ability is also used extensively as an analytic research tool (e.g., Cain et al., 1998; Cain & Oakhill, 1999; Nation et al., 1999; Nation & Snowling, 1998a; Nation & Snowling, 1998b; Oakhill, 1982; Oakhill et al., 1986; Stothard & Hulme, 1992; Stothard & Hulme, 1995; Yuill et al., 1989). The assessments it provides must therefore by accurate, and not subject to additional factors. However, in Experiment 6 it was found that performance

on the comprehension assessment provided by the Neale Analysis of Reading Ability was subject to question demands. The Neale Analysis of Reading Ability is also used to identify children with specific profiles across reading and comprehension ability. If reading and comprehension are indeed independent abilities, then selection of such groups is valid. The selection of these groups should be based on assessments of reading and comprehension that are themselves independent. Experiments 6 and 7 revealed that performance on the comprehension measure of the Neale Analysis of Reading Ability can depend on presentation condition, suggesting that this test does not provide independent assessments of reading and comprehension. The work presented in this chapter, therefore, investigates further both the accuracy of the separate measures offered by the Neale Analysis of Reading Ability, and the independence of those measures.

One concern regarding the assessment of reading accuracy provided by the Neale Analysis of Reading Ability, is that ability to provide the correct sound for each word (reading accuracy) does not necessarily reflect the ability to recognise and know the meanings of written words (word recognition). There can be situations in which ability to recognise and know the meanings of written words, and ability to say the words do not match. For example, some children – regardless of recognising and knowing the meaning of each word – rush quickly through the text, making a high number of careless mistakes. Other children take some time painstakingly piecing together the sounds of each word. It is therefore important to clarify what component of reading is targeted in research, and whether the assessment provided by the Neale Analysis of Reading Ability is suitable.

Similarly, work presented in this chapter addressed further whether ability to answer questions about text necessarily reflects internal comprehension of text. The Neale Analysis of Reading Ability assesses comprehension by assessing ability to answer questions that typically require descriptive and explicative responses, that fit the prescribed correct answer. However, the answers children provide rarely match exactly those ordained by the Neale Analysis of Reading

Ability to be correct, and answers then need to be assessed subjectively by the experimenter. Sometimes children parrot parts of text, and these responses can happen to match the correct answers, without the child appearing to have really understood the question or the text – i.e., children can answer questions through textual cues alone, and this behaviour might result in an overestimation of their comprehension ability. On the other hand, some children simply don't seem able to provide the answer, whilst their facial expressions and tone of voice while reading suggested good comprehension throughout the text. Sometimes children provide a response that is a feasible answer, and does reflect the text, but doesn't fit the predetermined correct response. Other children seem to lack the confidence to attempt any response. In Experiment 6 it was argued that the type of questions used in the Neale Analysis of Reading Ability make demands that are additional to comprehension, and evidence was found that these additional demands can result in an underestimation of text comprehension.

Finally, work presented in this chapter was conducted in order to question the independence typically assumed to exist between the reading accuracy and comprehension measures of the Neale Analysis of Reading Ability. When the Neale Analysis of Reading Ability is used to select children who exhibit specific profiles across reading accuracy and reading comprehension, it is assumed that the assessments provided by the two measures are independent of each other, so that individual performance can truly vary in both domains without performance in one domain affecting score on the other. However, the prescribed administration technique of the test confounds accuracy and comprehension abilities. It has been demonstrated that ability to read aloud words in context – as in the Neale Analysis of Reading Ability – can be dependent on comprehension skill (Nation & Snowling, 1998a; Nation & Snowling, 1998b. See Section 1.3.1 for details). It follows that the reading accuracy measure provided by the Neale Analysis of Reading Ability is not independent of reading comprehension ability. Furthermore, since children progress no further through the Neale Analysis of Reading Ability once they have made a prescribed number of reading errors, the number of questions over which comprehension ability is assessed is dependent on reading accuracy. This

concern can be illustrated by considering children with very poor reading accuracy, who reach the cut-off number of reading errors extremely early, and are therefore only assessed over very few comprehension questions. Even if such a child answered every question correct and clearly exhibited good comprehension ability, the child would obtain very low comprehension score. In short, whilst the Neale Analysis of Reading Ability offers separate standardised assessments of reading accuracy and of reading comprehension, doubts have arisen as to the independence of these measures.

The work conducted in Experiment 8 examined the validity of the reading accuracy measure provided by the Neale Analysis of Reading Ability, by comparing the performance of Year 3 children on this measure and on a word recognition test – a test with no vocalising and minimal comprehension demands. Similarly, the work conducted in Experiment 9 examined the validity of the comprehension measure provided by the Neale Analysis of Reading Ability, by comparing the performance of Year 3 children on this measure and a listening comprehension test – a test of comprehension that did not involve reading, and did not use the open-ended questions, demonstrated in Experiment 6 to be particularly difficult for some children.

The work conducted in Experiment 10 had two sections. The first investigated how children with specific profiles of performance across the reading accuracy and reading comprehension measures provided by the Neale Analysis of Reading Ability performed on alternative measures of reading and comprehension ability, provided by a word recognition test, and a listening comprehension test. This examined whether the Neale Analysis of Reading Ability is appropriate for selecting specific reading groups in research. Secondly, the independence of reading and comprehension was investigated – whether reading and comprehension abilities contribute independently to reading comprehension ability, and whether the assessments provided by the Neale Analysis of Reading Ability are independent of each other.

4.1 Development of the listening comprehension and unaided reading comprehension tests

For the purposes of Experiments 6, 9 and 10, assessments of listening comprehension and of unaided reading comprehension were required. Such tests did not already exist, and were therefore designed and developed as part of the work presented in this thesis. To provide comparison with comprehension performance on the Neale Analysis of Reading Ability, it was desirable that the two tests followed the same structure and text difficulty as the Neale Analysis of Reading Ability. Section 4.1 presents the development of these assessments.

4.1.1 Design of materials

For ease of administration, the listening and unaided reading comprehension assessments were presented as two components of one test. The original form of the test used the stories and pictures of the Neale Analysis of Reading Ability, Diagnostic tutor, (Neale, 1997), plus one additional story and accompanying set of questions composed to reflect the difficulty of Level 4 in the Neale Analysis of Reading Ability.

Stories were presented in order of increasing level, with a listening and then a silent reading story at each of Levels 1 to 4. Two further stories – one at Level 5 and one at Level 6 – were included for silent reading.

The original questions of the Diagnostic Tutor were turned into *true/false* statements. All question sets (except at Level 2) contained four memory questions and four questions that required an inference to be made; Level 1, with only four questions, had two of each question type. True and false statements were counterbalanced across memory and inference questions within each story.

The test was presented in booklet form. Each story appeared opposite its accompanying picture. The questions appeared on the following page, with the picture repeated.

The performance on the Neale Analysis of Reading Ability (form 2) of the 48 children (16 normal readers, less-skilled comprehenders and less-skilled decoders) who participated in an earlier study, revealed that 46 children successfully completed Level 4, whilst only 15 went beyond the accuracy cut-off to attempt Level 5. It was also found that performance on the comprehension questions of Levels 1 to 4 was sufficient to identify the different reading groups, with significantly lower performance across Levels 1 to 4 by less-skilled comprehenders than the normal readers and less-skilled decoders, $F(2,47)=47.760$, $p<0.001$. It was therefore decided that Levels 1 to 4 provided the optimum measure of comprehension ability, before becoming too difficult for most Year 3 readers. However, two extension stories were included to avoid particularly good readers obtaining ceiling scores; these were administered as silent reading stories, so that they could be attempted by skilled readers whilst less-skilled and slower readers completed Level 4.

Procedure

The test was administered by the experimenter to children in their usual classes. Each child was given a test booklet, and told to fill in their full names. The experimenter explained that she was interested in children's understanding of stories that they have heard or read themselves. Children were told that after each story they would hear or read some statements about that story. Some of the statements would be right, and some would be wrong, and the children's task was to say which, by marking with a tick or a cross.

Children were given a practice story, which was read aloud by the experimenter, and the class and experimenter then worked through the practice questions together.

The listening stories were read aloud by the experimenter, while children followed the texts in their own booklets. They then turned the page, and each question was read aloud to them whilst they followed it in their booklets. They then decided individually whether to tick or cross each statement. Children worked in strict independence. Referring back to the story was not permitted.

Children read the silent reading stories in their own time, independently. They were informed that some of the words would be difficult, and that they must do their best to understand what was meant. When they had finished reading the story they turned the page to answer the questions in their own time. Children were not allowed to turn back to the story whilst answering questions, and were not permitted to turn onto the next story before everyone else. To assure this, children were told to turn their booklets over once they had finished the questions.

After the Level 4 listening story, children were told that they had finished the listening stories, and now had three silent reading stories that they could work through in their own time, turning on to the next story upon finishing the questions. They were told that all children should try to complete the level 4 story, and anyone who finished early should try the extension stories. They were warned that the last two stories were designed to be difficult, and that not everyone would find them within their ability.

Scoring

Performance on each component of the test was scored as number of items correct.

4.1.2 Development of materials

Pilot stage 1

Initial piloting on a class of 23 Year 3 children suggested that the stories did not increase smoothly in difficulty with level. Table 4.1 shows mean scores on each story. To verify this, a second pilot stage administered the whole test to some children in listening form, and to other children entirely in silent reading form.

Table 4.1. Scores from pilot stage 1 of comprehension test development, and number of children attempting each story.

Level	Condition	Story	Maximum score	Mean	SD	N
1	Listening	Dog	4	3.8	0.9	23
1	Reading	Box	4	3.5	1.3	21
2	Listening	Helicopter	8	5.5	3.7	23
2	Reading	Fishing	8	6.2	3.4	21
3	Listening	Penguins	8	4.1	4.0	23
3	Reading	Seagull	8	5.4	3.8	21
4	Listening	Wizard	8	6.1	3.8	22
4	Reading	Ghosts	8	5.1	3.9	21
5	Reading	Submarine	8	5.0	3.9	20
6	Reading	Volcano	8	4.8	3.9	17

Pilot stage 2

To examine the equivalence of the two story sets, Year 3 children were administered both components of the test, either listening to both (32 children), or reading both (24 children). Table 4.2 shows mean scores on each story in each condition. Inspection of the results suggested that there was some imbalance between the sets at Levels 2 and 3. To investigate the equivalence of the story sets, a series of one-way analyses of variance, with story set as factor, were conducted on accuracy scores obtained at each of Levels 1 to 4, in listening and reading conditions separately. The analyses revealed non-significant main effects of story set at Level 1, both when listening, $F(1,6)<1$, and when reading, $F(1,6)=1.12$, $p>.05$, and at Level 2, when listening, $F(1,14)=2.82$, and when reading, $F(1,14)=1.97$, both $p>.05$. The analyses revealed non-significant main effect of story set at Level 3 when listening, $F(1,14)=2.77$, $p>.05$, and significant main effect of story set at Level 3 when reading, $F(1,14)=5.79$, $p<.05$, with more correct answers provided after the *Seagull* story than after the *Penguin* story. The analyses revealed non-significant main effects of story set at Level 4, when listening, $F(1,14)=1.74$, $p>.05$, and when reading, $F(1,14)<1$.

Table 4.2. Scores from pilot stage 2 of comprehension test development.

Maximum score at Level 1 was 4, maximum score at Levels 2 to 4 was 8.

			Listening			Silent reading		
Level	Story	Maximum score	Mean	SD	N	Mean	SD	N
1	Dog	4	3.6	1.2	31	3.5	1.3	24
1	Box	4	3.6	1.2	31	3.2	1.6	24
2	Helicopter	8	5.9	3.5	31	6.1	3.4	24
2	Fishing	8	7.2	2.4	31	7.0	2.6	24
3	Penguins	8	4.4	4.0	31	4.3	4.0	24
3	Seagull	8	5.6	3.7	31	5.8	3.6	24
4	Wizard	8	6.4	3.2	31	5.7	3.6	21
4	Ghosts	8	5.7	3.7	31	5.7	3.6	21
5	Submarine	8	5.6	3.7	31	4.6	4.0	11
6	Volcano	8	4.6	4.0	31	3.8	4.0	10

To investigate the source of imbalance between the story sets, a series of Pearson correlations, by child, was conducted between the accuracy of each item and the total accuracy of each story set. The analyses identified two low correlating items, both in story set A at Level 3 (*Penguins* story). These two items were correctly marked with accuracy below chance across all children, in reading and listening presentation conditions. One other item was marked with low accuracy, in Level 2, story set A (*Helicopter*). These three items were removed from the data, and the analyses repeated. Inspection of the new means revealed clear improvement in the balance of the two sets across the levels. The three-way analysis of variance with level, story set, and condition with factors, revealed non-significant interaction between level and set, $F(3,45)=1.75$, $p>.05$. The t-test comparisons between accuracies of the two story sets at each level, in listening and reading conditions separately, revealed reduced difference between the stories at Levels 2 and 3.

In summary, although there were few significant differences between the stories at each level, individual questions were shown to be especially difficult for the

children, and removal of these items was found to improve the equivalence of the two story sets. The next step, therefore, was to replace these items.

The Level 2 item *The flashing lights might have been fire engines* tapped knowledge that was not clearly prompted by the story. This was changed to *The helicopter was on fire.* The two Level 3 items were especially obscure, relying on picking up the single word cues *staggering* to demonstrate the difficulty of the journey, and *well-worn* to show that the journey was frequent. The sentences within the story were re-written to make this information more salient: *Once more, they struggled up the sand-hills to their burrows.* In addition to this change, the statement *The penguins make this journey often* was altered to *The penguins have made this journey before.*

Pilot stage 3

The new test version was administered to two new Year 3 classes – one class of 21 listening to all stories and one class of 23 reading all stories. The three-way analysis of variance of the accuracy of items of Levels 1 to 4, with level, set and condition as factors, revealed significant main effect of level, $F(3,48)=14.53$, $p<.001$, non-significant main effects of condition, $F(1,48)<1$, and set, $F(1,48)=1.80$, $p>.05$ non-significant interaction between condition and level, $F(3,48)=1.83$, $p>.05$, non-significant interaction between condition and set, $F(1,48)=1.18$, $p>.05$, and non-significant interaction between condition, set and level, $F(3,48)=2.19$, $p>.05$. However, the analysis did reveal significant interaction between level and set, $F(3,48)=5.81$, $p<.05$, indicating that at some level the stories were not equivalent. Inspection of the results suggested that the problem remained within Level 3, with consistently higher performance on the *Seagull* story (mean listening score 5.54, standard deviation 0.90; mean reading score 5.83, standard deviation 0.71, maximum score 8) than on the *Penguin* story (mean listening score 3.83, standard deviation 1.25; mean reading score 4.28, standard deviation 0.33, maximum score 8). This was confirmed, by repeating the above analysis of variance, with the level 3 stories removed; the analysis now revealed non-significant interaction between level and set, $F(2,34)=2.32$, $p>.05$.

Pilot stage 4

To resolve the ongoing problems with Level 3, two new stories and question sets were written, estimated to match the level of difficulty and vocabulary of the *Seagull* story and the Level 3 stories in forms 1 and 2 of the Neale Analysis of Reading Ability. These new stories were administered to the same Year 3 classes who participated in pilot stage 3, in the same test conditions as stage 3.

The two-way analysis of variance on the accuracy with which each item of the new Level 3 stories was marked, with condition and story as factors, revealed non-significant main effect of condition, $\underline{F}(1,14)=4.67$, $\underline{p}>.05$, and non-significant main effect of story, $\underline{F}(1,14)<1$, and non-significant interaction between condition and story, $\underline{F}(1,14)<1$.

The new Level 3 stories were substituted into the data from stage 3, to give item accuracy of the final test version. Questions that were not attempted scored zero. Table 4.3 shows mean scores in each condition on each story. The three-way analysis of variance of item accuracy across the first four levels, with level, set and condition as factors, revealed significant main effect of level, $\underline{F}(3,48)=4.59$, $\underline{p}<.01$, significant main effect of condition, with higher accuracy when listening than reading, $\underline{F}(1,48)=4.48$, $\underline{p}<.05$, non-significant main effect of set, $\underline{F}(1,48)<1$, and non-significant interaction between level and set, $\underline{F}(3,48)=1.93$, $\underline{p}>.05$.

This was the final version of the test. The listening comprehension component of the test was comprised of story set A, and the unaided reading comprehension component was constructed of story set B. Stories were presented in the order shown in Table 4.3, alternating between listening and unaided reading.

Table 4.3. Mean number of questions correct in pilot stages 3 and 4 of comprehension test development, standard deviations in parentheses.

Level	Set	Story	Maximum score	Condition	
				Listening (N=21)	Reading (N=23)
1	A	Dog	4	3.8 (0.5)	3.6 (0.9)
1	B	Box	4	3.6 (0.6)	3.2 (1.0)
2	A	Helicopter	8	5.6 (1.5)	6.1 (1.7)
2	B	Fishing	8	7.1 (0.8)	6.2 (1.6)
3	A	Camping	8	6.4 (1.5)	6.0 (1.9)
3	B	Fire	8	6.8 (1.3)	6.2 (2.0)
4	A	Wizard	8	6.1 (1.7)	5.7 (1.8)
4	B	Ghosts	8	5.4 (1.4)	5.5 (1.6)
5	B	Submarine	8	3.8 (2.4)	4.1 (2.3)
6	B	Volcano	8	3.1 (2.3)	3.2 (2.1)

4.2 Experiment 8

The first aim of the work presented in this chapter was to examine whether some aspects of the Neale Analysis of Reading Ability might result in an underestimation of some children's ability. In Experiment 8, a large sample was used to explore whether some children have difficulty with the vocalisation aspect and concurrent comprehension demands of the reading accuracy component of the Neale Analysis of Reading Ability, despite sufficient word recognition. This would predict that all who score highly on reading accuracy would also score highly on word recognition, that all who score poorly on word recognition would also score poorly on reading accuracy, and that some would score poorly on reading accuracy despite scoring well on word recognition. If the discrepancy of the latter subgroup simply reflected fluctuations in performance for reasons other than the proposed differences between the two tests, the converse subgroup should also appear – obtaining high scores on reading accuracy and low scores on word recognition.

4.2.1 Method

Participants

A total of two-hundred and twenty-one Year 3 children from four Bristol schools participated in this study. Whole classes, and where applicable, all classes in the year group, were tested, to ensure that the natural range of abilities was included. The sample ranged in age from 7.17 years to 8.50 years, with mean age 7.90 years, and standard deviation 0.30 years.

Materials

All children were administered the Word Decision Test of the Reading Decision Test, form A, (Baddeley, Gathercole, & Spooner, in press), and the Neale Analysis of Reading Ability (Neale, 1997). The Word Decision Test yields scores for ability to distinguish single written words from single written non-words. The Neale Analysis of Reading Ability yields standardised reading ages for decoding accuracy and comprehension.

Procedure

Children were first administered the Word Decision Test, in their usual classes. The Neale Analysis of Reading Ability was then administered to each child individually, over the few days following the Word Decision Test administration.

4.2.2 Results

Performance on the Word Decision Test was measured as number of items correct. The mean score, from the standardisation data, for Year 3 is 33/40 for the Word Decision Test, with standard deviation 8.12. The sample who participated in Experiment 8 obtained mean score of 36.37 (standard deviation 6.49) on the Word Decision Test, and mean reading accuracy age on the Neale Analysis of Reading Ability of 8.89 years (standard deviation 1.93 years), and mean reading comprehension age of 8.14 years (standard deviation 1.47 years).

The first analysis investigated the correspondence between performance on the Word Decision Test and the reading accuracy measure provided by the Neale Analysis of Reading Ability. Pearson correlations were conducted between raw scores obtained on the reading accuracy measure of the Neale Analysis of Reading Ability and scores obtained on the Word Decision Test, for children who were administered form 1 of the Neale Analysis of Reading Ability and those administered form 2 separately. The analyses revealed significant correlation between performance on the Word Decision Test and reading accuracy measure of the Neale Analysis of Reading Ability, for both children who were administered form 1 of the Neale Analysis of Reading Ability, $r=.72$, $n=96$, and children who were administered form 2, $r=.63$, $n=125$, both $p<.001$. The correlation coefficients were higher for children administered form 1 than those administered form 2. The children who were administered form 2 attended a high performance school, and generally performed better than those administered form 1. Indeed, 54 of the 125 children who were assessed on form 2 of the Neale Analysis of Reading Ability scored 40/40 on the Word Decision Test, compared to 22 of the 96 children who were assessed on form 1 of the Neale Analysis of Reading Ability. To verify whether the lower correlation coefficient for form 2 children resulted from higher proportions of ceiling scores, the Pearson correlations between raw accuracy score on the Neale Analysis of Reading Ability and corrected score on the Word Decision Test were repeated, with children who obtained 40/40 on the Word Decision Test omitted. The analyses revealed significant correlation between performance on the two tests, both for children who were assessed on form 1 of the Neale Analysis of Reading Ability, $r=.71$, $n=74$, and those assessed on form 2, $r=.62$, $n=71$, both $p<.001$. Correlation coefficients were still higher for children assessed on form 1 of the Neale Analysis of Reading Ability than those assessed on form 2, suggesting that the two versions of the Neale Analysis of Reading Ability might not be perfectly equivalent. Nonetheless, the Pearson correlations for both groups of children revealed that performance on the Word Decision Test corresponded with performance on the reading accuracy measure of the Neale Analysis of Reading Ability.

To examine the correspondence between performance on the Word Decision Test and the reading accuracy measure of the Neale Analysis of Reading Ability beyond age, the Pearson correlations between raw accuracy score and corrected score on the Word Decision Test were repeated with age partialled out, for children who were administered form 1 of the Neale Analysis of Reading Ability and those administered form 2 separately. The analyses revealed significant correlations between performance on the Word Decision Test and the reading accuracy measure of the Neale Analysis of Reading Ability beyond age, in children who were assessed on form 1 of the Neale Analysis of Reading Ability, $r=.73$, $n=96$, and those assessed on form 2, $r=.63$, $n=125$, both $p<.001$. These correlations were repeated with children who scored 40/40 on the Word Decision Test omitted, and confirmed the significant correlation, in children assessed on form 1 of the Neale Analysis of Reading Ability, $r=.71$, $n=74$, and those assessed on form 2, $r=.62$, $n=71$, both $p<.001$.

The next stage of analysis investigated the consistency of individual performance across the two tests. Performance on the reading accuracy measure of the Neale Analysis of Reading Ability was categorised as high if reading accuracy age was at least equivalent to chronological age, as low if reading accuracy age was at 12 months or more below chronological age, and as middle performance if reading accuracy age fell between these two markers. Word recognition performance was categorised as high if above the population mean for Year 3 (scores of 33/40 and more), as low if score was more than one standard deviation below the mean for Year 3 (scores of 25/40 and less), and categorised as middle performance if Word Decision Test score fell between 26 and 32 out of 40. Table 4.4 presents the contingency table of performance across the two tests.

Table 4.4 demonstrated that nearly all children who scored well on reading accuracy also scored well on word recognition, and that most children with low word recognition ability also scored poorly on reading accuracy. A number of children obtained low or middle reading accuracy scores despite high word recognition scores, and very few children exhibited the converse profile. This

pattern of results is consistent with the prediction that reading accuracy demands more than word recognition. Cross-tabulation analysis was conducted to investigate whether category of performance on one test reliably predicted category of performance on the other, and revealed both that word recognition predicted reading accuracy, lambda=.29, d.f.=2, p<.001, and that reading accuracy predicted word recognition, lambda=.35, d.f.=2, p<.05.

Table 4.4. Contingency table of scores in Experiment 8.

	Word recognition		
Reading accuracy	High	Middle	Low
High	152	1	0
Middle	33	4	3
Low	7	4	17

In summary, Experiment 8 revealed good correspondence between the measure of reading accuracy provided by the Neale Analysis of Reading Ability and the measure of word recognition provided by the Word Decision Test, in the general population. Good reliability in performance category was also demonstrated across the two tests.

4.2.3 Discussion

Experiment 8 was conducted to explore whether the reading accuracy measure of the Neale Analysis of Reading Ability might underestimate the reading ability of some children. Reading accuracy performance was compared to word recognition performance on the Word Decision Test, a test with no vocalising and minimal comprehension demands. It was proposed that if the vocalising and comprehension demands present in the Neale Analysis of Reading Ability proved problematic for some children, a subgroup of children would be found who obtained low reading accuracy scores despite good word recognition, in the absence of the converse subgroup.

In Experiment 8, there was some observational evidence that the crucial subgroup might exist, in the absence of the converse subgroup, but analysis

demonstrated that performance on each test was reliably predicted by performance on the other. This evidence suggests that, while some children might find the vocalising and comprehension demands of the reading accuracy measure difficult, the risk of underestimating reading ability with the Neale Analysis of Reading Ability is not significant.

4.3 Experiment 9

Experiment 9 was conducted to investigate whether the reading comprehension measure provided by the Neale Analysis of Reading Ability might underestimate comprehension ability, relative to a listening comprehension assessment with no reading and reduced question demands. This would predict that all who score highly on the reading comprehension measure of the Neale Analysis of Reading Ability would also score highly on listening comprehension, that all who score poorly on listening comprehension would also score poorly on reading comprehension, and that some would score poorly on reading comprehension despite scoring well on listening comprehension. If this latter subgroup merely reflected normal fluctuations in test performance, the converse subgroup would also appear – obtaining high reading comprehension and low listening comprehension scores.

4.3.1 Method

Participants

A total of two-hundred and twelve Year 3 children from six urban Bristol schools participated in this study. Whole classes were tested, to include the natural variety of abilities. The sample ranged in age from 7.17 years to 8.58 years, with mean age 7.97 years, and standard deviation 0.32 years.

Materials

All children were administered the listening comprehension test, described in Section 4.1, and form 1 of the Neale Analysis of Reading Ability (Neale, 1997), which yielded standardised reading ages for accuracy and comprehension.

Procedure

Children were first administered the listening comprehension test, in their usual classes, according to the procedure described in Section 4.1. The Neale Analysis of Reading Ability was then administered to each child individually, over the few days following administration of the listening comprehension test.

4.3.2 Results

The sample obtained mean score of 23.66 (maximum score 28, standard deviation 2.86) on the listening comprehension test, and mean reading accuracy age of 8.49 years (standard deviation 1.59 years) on the Neale Analysis of Reading Ability, and mean reading comprehension age of 7.91 years (standard deviation 1.36 years) on the Neale Analysis of Reading Ability.

To investigate the correspondence between the measures of comprehension ability provided by the Neale Analysis of Reading Ability and by the listening comprehension test, a Pearson correlation was conducted between reading comprehension raw score obtained on the Neale Analysis of Reading Ability and listening comprehension score. The analysis revealed significant correlation between reading comprehension and listening comprehension, $r=.58$, $n=212$, $p<.001$.

To investigate the correspondence between the measures of reading comprehension and listening comprehension, beyond age and reading accuracy, a Pearson correlation was conducted between reading comprehension raw score obtained on the Neale Analysis of Reading Ability and listening comprehension score, with age and accuracy raw score obtained on the Neale Analysis of Reading Ability partialled out. The analysis revealed significant correlation between reading comprehension and listening comprehension, when age and reading accuracy were controlled for, $r=.53$, $n=212$, $p<.001$.

The next analysis investigated reliability of performance category across the two comprehension assessments. Reading comprehension performance was categorised as high if reading comprehension age on the Neale Analysis of

Reading Ability was equivalent to or above chronological age, as low if reading comprehension age was 12 months or more below chronological age, and categorised as middle performance if reading comprehension age fell between these two markers. Since the assessment of listening comprehension was not a standardised test, individual performance was categorised relative to performance of the Experiment 9 sample. Listening comprehension was categorised as high if higher than the mean (a score of 24/28 or more), as low if more than one standard deviation below the mean (a score of 20/28 or less), and categorised as middle performance if scoring between these two cut-offs. The contingencies of performance category across the two tests are presented in Table 4.5.

Table 4.5. Contingency table of scores in Experiment 9.

	Listening comprehension		
Reading comprehension	High	Middle	Low
High	70	22	1
Middle	34	24	4
Low	18	19	20

The contingency table (Table 4.5) demonstrated that most children who obtained high reading comprehension score also obtained high listening comprehension score, and that nearly all children who exhibited low listening comprehension also scored poorly on reading comprehension. However, the children who obtained low reading comprehension scores appeared equivalently distributed across the bands of performance on listening comprehension. A cross-tabulation analysis was conducted to explore the predictability of performance between the tests. The analysis revealed that, while listening comprehension was a significant predictor of reading comprehension, lambda=.17, d.f.=2, $p<.01$, reading comprehension was not a significant predictor of listening comprehension, lambda=.02, d.f.=2, $p>.05$.

To examine whether poor reading ability might account for the occurrence of children with poor reading comprehension despite good listening

comprehension, the reading accuracy scores obtained by these children on the Neale Analysis of Reading Ability were inspected. Of the 18 children who obtained reading comprehension age 12 months or more below chronological age despite above average listening comprehension, 11 had exhibited reading accuracy below what was expected of their age, of whom 9 obtained reading accuracy age at least 6 months below chronological age. However, of the subgroup, only two children exhibited reading accuracy age more than 12 months below chronological age. The mean discrepancy between accuracy age and chronological age for this subgroup was 0.09 years below chronological age (standard deviation 1.53 years). To ascertain whether this subgroup reliably exhibited reading accuracy below what was expected of their age, a one-sample t-test analysis was conducted on the discrepancy between accuracy age and chronological age within this subgroup, with test value zero. The analysis revealed that this group did not significantly obtain reading accuracy age lower than chronological age, $t(17)=0.34$, $p>.05$.

In summary, Experiment 9 revealed good correspondence between reading comprehension as assessed by the Neale Analysis of Reading Ability and listening comprehension, in the general population. However, a number of children who exhibited poor performance on the Neale Analysis of Reading Ability scored well on the listening comprehension test, in the absence of the converse subgroup, and reading comprehension performance did not reliably predict listening comprehension score. In addition, while some children who obtained low reading comprehension despite high listening comprehension score exhibited poor reading accuracy, this subgroup were not generally poor readers.

4.3.3 Discussion

Experiment 9 was conducted to investigate whether the reading comprehension assessment provided by the Neale Analysis of Reading Ability risked underestimating comprehension ability, due to the additional reading and question demands made by the test. Reading comprehension performance was compared to performance on a listening comprehension test – a task with no

reading and reduced question demands. It was proposed that if the demands in the reading comprehension assessment impaired the performance of some children, a subgroup would appear with poor reading comprehension despite good listening comprehension, in the absence of the converse subgroup.

It was found that a subgroup did indeed exist, obtaining low reading comprehension score despite good listening comprehension. Furthermore, it was found that while listening comprehension predicted reading comprehension, reading comprehension did not predict listening comprehension. These findings support the hypothesis that the Neale Analysis of Reading Ability makes demands that are additional to comprehension, and thereby risks underestimating comprehension ability in individuals for whom those demands present some difficulty.

One demand that the Neale Analysis of Reading Ability makes beyond comprehension per se is the requirement to read aloud. It is therefore possible that children with poor reading accuracy may not have derived all the information from text necessary for comprehension. Although reading errors were corrected during administration of the Neale Analysis of Reading Ability, children with poor reading accuracy may have found the interruptions disruptive and the necessary alterations to their representations confusing. The quality of information derived from text during slow, disjointed and interrupted reading may never match the quality of information derived during smooth, focused reading. Secondly, while children were expected to distribute effort between reading accuracy and reading comprehension, children with poor reading ability may have concentrated effort into reading and therefore away from comprehension. It is therefore possible that children with poor reading ability were impaired in their comprehension of the Neale Analysis of Reading Ability because of the demands made by the reading component.

In addition, the prescribed administration technique of the Neale Analysis of Reading Ability risks confounding accuracy and comprehension. Children progressed no further through the test once they made a prescribed number of

reading errors, and, therefore, the number of questions over which comprehension ability was assessed was dependent on reading accuracy. This concern can be illustrated by considering children with very poor reading accuracy, who reached the cut-off number of reading errors extremely early, and were therefore only assessed over very few comprehension questions. Even when such a child answered every question correctly and clearly exhibited good comprehension ability, the child obtained very low comprehension score.

The listening comprehension test, on the other hand, made no reading demand, and therefore removed the risk of handicapping poor readers. Evidence from Experiment 9 suggests that this did account for the low reading comprehension of some children with good listening comprehension, but the subgroup whose comprehension was underestimated by the Neale Analysis of Reading Ability were not generally poor readers.

Other demands made by the Neale Analysis of Reading Ability beyond comprehension per se are those placed by the open-ended questions. As discussed in Experiment 6, the Neale Analysis of Reading Ability asked complex questions that required the formulation of answers that described and explained what the child understood. It is therefore possible that some children obtained a lower score on the Neale Analysis of Reading Ability than their actual comprehension ability merited due to poor expressive speech, or low confidence at attempting answers. The listening comprehension test used in Experiment 9 removed these potential hurdles by requiring a forced-choice response, and thereby no longer depended on expressive speech. It was demonstrated in Experiment 6 that groups of children with highly separated comprehension scores on the conventional version of the Neale Analysis of Reading Ability, did not continue to differ in a similar test when comprehension was assessed by forced-choice questions. Experiment 9, therefore, adds further evidence to the earlier finding that the open-ended questions used in the Neale Analysis of Reading Ability can underestimate comprehension ability, and risk including in less-skilled comprehension groups children whose difficulties instead lie in some aspect of formulating expressive answers.

4.4 Experiment 10

The first aim of Experiment 10 was to investigate whether the Neale Analysis of Reading Ability could reliably identify specific reading and comprehension profiles. Groups were selected according to their performance on the Neale Analysis of Reading Ability, and their performance across alternative assessments of reading and comprehension was investigated. The finding that some children within specific groups, as identified by the Neale Analysis of Reading Ability, exhibit different profiles of performance across other assessments of reading and comprehension, would suggest that the Neale Analysis of Reading Ability might not be able to provide accurate group selection. This would have important implications for the use of the Neale Analysis of Reading Ability in experimental research that relies on group comparisons.

The second aim of the work conducted in Experiment 10 was to investigate the independence of reading and comprehension – as abilities, and as components of the Neale Analysis of Reading Ability. First, it was investigated whether ability to recognise written words and ability to comprehend were independent. The finding that reading and comprehension were indeed independent abilities would justify the selection of experimental groups who differ in profile across these abilities. The independence of these two skills was ascertained by examining their contribution to performance on an unaided reading comprehension test – a task which relied on both ability to recognise written words, and ability to derive meaning. The finding that performance on the unaided reading comprehension test was predicted by both word recognition and comprehension ability, and by these skills independent of each other, would indicate that reading and comprehension are indeed independent abilities.

Next, the independence of the measures offered by the Neale Analysis of Reading Ability was investigated. If reading and comprehension were demonstrated to be independent abilities, the reading and comprehension components of the Neale Analysis of Reading Ability must also be independent, if this test is to be used as an analytic tool of assessment. In the

final section of Experiment 10, therefore, the contributions of the reading and comprehension measures provided by the Neale Analysis of Reading Ability to performance on the unaided reading comprehension test were investigated.

4.4.1 Method

Participants

A total of 92 Year 3 children participated in both Experiment 8 and Experiment 9, completing the Word Decision Test, the listening comprehension test, and the unaided reading comprehension test. This sample ranged in age from 7.17 years and to 8.33 years, with mean age 7.98 years, and standard deviation 0.29 years.

Materials

All children were administered the listening comprehension and unaided reading comprehension tests, described in Section 4.1, the Word Decision Test, form A, (Baddeley et al., in press), and form 1 of the Neale Analysis of Reading Ability (Neale, 1997).

Procedure

Children were first given the Word Decision Test, the listening comprehension test and the unaided reading comprehension test in one sitting in their usual classes. The Neale Analysis of Reading Ability was then administered to each child individually, over the few days following the administration of the class tests.

4.4.2 Results

The performance of the sample on the measures provided by the Neale Analysis of Reading Ability, the Word Decision Test, the listening comprehension test and the unaided reading comprehension test is summarised in Table 4.6.

Table 4.6. Mean scores obtained by the sample of children in Experiment 10. Ages in years. Standard deviations in parentheses.

Neale Analysis of Reading Ability Accuracy Age	8.3 (19.8)
Neale Analysis of Reading Ability Comprehension Age	7.8 (16.5)
Word Decision Test (maximum = 40, population mean = 33)	34.7 (8.3)
Listening comprehension (maximum = 28, Experiment 9 mean = 23.7)	23.7 (3.3)
Unaided reading comprehension (maximum = 44, Experiment 9 mean = 29.7)	29.0 (6.2)

The first stage of analysis investigated how children with specific profiles across the reading measures of the Neale Analysis of Reading Ability scored on the Word Decision Test and the listening comprehension test. Several groups were selected on the basis of performance on the Neale Analysis of Reading Ability. Children who scored at least age-appropriately on both the accuracy and comprehension measures of the Neale Analysis of Reading Ability were identified as skilled readers. Two groups of less-skilled comprehenders, and two groups of less-skilled decoders, were selected in the following ways.

Children who obtained at least age-appropriate accuracy, but some deficit of comprehension below chronological age on the Neale Analysis of Reading Ability, were identified as less-skilled comprehenders. In this study two degrees of less-skilled comprehension were considered: those with a comprehension deficit of at least six months below age, and a more extreme group whose deficit was at least twelve months. The first group was larger, but a concern arose regarding the reliability of this six-month deficit, as it can arise from just two questions answered incorrectly. The more extreme group was therefore included to check on the reliability and generality of the findings with the less severe reading comprehension group. The severe group was included within the less severe group.

Two groups of less-skilled decoders were selected, both with reading accuracy age on the Neale Analysis of Reading Ability six months or more below

chronological age, but differing in comprehension age on the Neale Analysis of Reading Ability. It is typically required that less-skilled decoders exhibit at least age-appropriate comprehension performance on the Neale Analysis of Reading Ability, but children are limited by their reading accuracy in the number of comprehension questions they can attempt. This can result in less-skilled decoders being unable to obtain age-appropriate comprehension, not through poor comprehension, but through poor reading. While it is desirable to ascertain that less-skilled decoders are good comprehenders, this is not always possible with the Neale Analysis of Reading Ability, and specifying age-appropriate comprehension can exclude many genuine less-skilled decoders from the sample.

In this study, two groups of less-skilled decoders were investigated, both with the same criteria of reading accuracy, but differing in minimum comprehension ability. A liberal group included children whose accuracy age on the Neale Analysis of Reading Ability was six months or more below their chronological age, and comprehension age no more than twelve months below their chronological age. This group provided maximum numbers for analysis, permitting the inclusion of children whose comprehension score might have been impaired by their poor decoding. The conservative group included only those children who exhibited an accuracy deficit of at least six months, and comprehension age no more than six months below their chronological age. This conservative group was included to check the reliability and generality of the more liberal less-skilled decoder group, in children whose poor performance was more specific to reading accuracy.

Groups were selected in this manner, and their mean scores on the Word Decision Test and the listening comprehension test were inspected. Group performances are summarised in Table 4.7.

Table 4.7. Mean performance of groups as selected by the Neale Analysis of Reading Ability in Experiment 10. Ages in years. Standard deviations in parentheses.

	Neale selection criteria				Neale Analysis of Reading Ability reading ages		Word Decision	Listening comprehension
Groups	Accuracy	Comprehension	N	Age	Accuracy	Comprehension	Test	test
Skilled readers	Age or above	Age or above	34	7.8 (0.3)	9.6 (1.4)	9.2 (1.1)	39.0 (1.3)	25.2 (2.3)
less-skilled comprehenders	Age or above	At least 6m below age	14	7.7 (0.3)	8.7 (1.0)	6.5 (0.5)	38.4 (1.5)	21.7 (3.4)
	Age or above	At least 12m below age	10	7.8 (0.3)	8.8 (1.1)	6.4 (0.5)	38.1 (1.5)	21.0 (3.8)
less-skilled decoders	At least 6m below age	12m below age or better	15	7.8 (0.3)	6.7 (0.5)	7.4 (0.6)	27.3 (8.3)	23.7 (3.5)
	At least 6m below age	6m below age or better	8	7.9 (0.4)	7.0 (0.5)	7.8 (0.6)	32.1 (7.6)	23.3 (4.4)

Table 4.8. Number of children in each reading group achieving three profiles across the Word Decision Test and the listening comprehension test, in Experiment 10. Also expressed as percentage of each group.

Groups	Neale selection criteria		Word Decision Test ≥33/40 and listening comprehension ≥24/28	Word Decision Test ≥33/40 and listening comprehension <24/28	Word Decision Test ≤ 29/40 and listening comprehension ≥24/28
	Accuracy	Comprehension			
Skilled readers	Age or above	Age or above	25 (74%)	9 (26%)	0 (0%)
Less-skilled comprehenders	Age or above	At least 6m below age	6 (43%)	8 (57%)	0 (0%)
	Age or above	At least 12m below age	4 (40%)	6 (60%)	0 (0%)
Less-skilled decoders	At least 6m below age	12m below age or better	5 (33%)	0 (0%)	5 (33%)
	At least 6m below age	6m below age or better	5 (63%)	0 (0%)	1 (13%)

The groups of skilled readers and less-skilled comprehenders obtained higher mean scores on the Word Decision Test than the groups of less-skilled decoders. Conversely, the groups of skilled readers and less-skilled decoders obtained higher mean scores on the listening comprehension test than the groups of less-skilled comprehenders.

The next step was to ascertain what proportion of the reading groups, as identified by the Neale Analysis of Reading Ability, fell into certain profiles of performance on the Word Decision Test and the listening comprehension test. At least age-appropriate word recognition ability was indicated by the Word Decision Test by a score of 33/40 and above, and low reading ability by a score of less than 33/40. Experiment 9 established that at least age-appropriate comprehension ability corresponded with a score of 24/28 and above on the listening comprehension test, and low comprehension ability by less than 24/28. Three score-profiles of interest result, that correspond in theory to the three reading groups: good performance on both the Word Decision Test and listening comprehension test, corresponding to skilled reading; good performance on the Word Decision Test but low performance on listening comprehension, corresponding to less-skilled comprehension; and low performance on the Word Decision Test but good performance on listening comprehension, corresponding to less-skilled decoding. The number of each reading group as identified by the Neale Analysis of Reading Ability performing within the above three profiles was derived, and shown in Table 4.8. These numbers were also expressed as percentages of each group. Skilled readers – as selected by performance on the Neale Analysis of Reading Ability – tended to score above average on both the Word Decision Test and the listening comprehension test. Less-skilled comprehenders – as selected by performance on the Neale Analysis of Reading Ability – tended to score above average on the Word Decision Test, and below average on the listening comprehension test, but many scored above average on both tests. Less-skilled decoders – as selected by performance on the Neale Analysis of Reading Ability – tended to score above average on listening comprehension, and

exhibited equivalent tendency to score above or below average on the Word Decision Test.

The next analyses investigated whether performance across the reading and comprehension measures of the Neale Analysis of Reading Ability could reliably predict specific profiles of performance on the Word Decision Test and the listening comprehension test. Group categorisation was determined according to the following criteria on performance on the Word Decision Test and the listening comprehension test: Skilled reading was indicated by at least age-appropriate performance on the Word Decision Test, 33/40 or more, and listening comprehension test, 24/28 or more; Less-skilled comprehension was indicated by at least age-appropriate performance on the Word Decision Test, 33/40 or more, and low performance on the listening comprehension test, below 24/28; Less-skilled decoding was indicated by low performance on the Word Decision Test, below 33/40, and good performance on the listening comprehension test, 24/28 or more.

A discriminant function analysis, with raw scores obtained on the reading accuracy and reading comprehension measures of the Neale Analysis of Reading Ability as predictors, was conducted on the three categories of reading profiles exhibited on the Word Decision Test and the listening comprehension test as described above. The analysis yielded two discriminant functions, with a significantly reliable relationship between groups and predictors, Chi-squared (4) = 28.72, $p<.001$ (Wilks' Lambda = .68). Function 1 accounted for 59.0% of the variance, with loadings of 1.18 and –1.02 on the decoding and comprehension scores of the Neale Analysis of Reading Ability respectively. Function 2 was also a significantly reliable predictor, Chi-squared (1) = 12.01, $p=.001$ (Wilks' Lambda = .85). Function 2 accounted for 41.0% of the variance, with loadings of 0.40 and 0.71 on the decoding and comprehension scores of the Neale Analysis of Reading Ability respectively. The discriminant functions correctly assigned 65% of cases to their original groups, correctly recognising 80% of children who scored well on the Word Decision Test and the listening comprehension test, but only 46% of children who scored well on the Word

Decision Test and poorly on the listening comprehension test, and none of those who scored well on the listening comprehension test and poorly on the Word Decision Test.

The next set of analyses investigated whether word recognition ability and comprehension ability were independent contributors to unaided reading comprehension proficiency. The Word Decision Test provided a measure of word recognition, and the listening and reading comprehension components of the class test provided measures of listening and unaided reading comprehension. These three measures were obtained independently. To examine whether word recognition and comprehension ability interact to provide further assistance with reading comprehension, an additional predictor variable was derived by multiplying scores obtained on the Word Decision Test and the listening comprehension test. To ascertain the degree to which reading comprehension is aided by the interaction between word recognition and comprehension ability beyond the two abilities independently, this variable was entered into the model at the last step. Since the interaction variable was derived from the other predictors, the variables do not satisfy the independence requirement for linear models. A non-linear regression model was therefore applied.

Two non-linear hierarchical multiple regression analyses were conducted, with unaided reading comprehension as the dependent variable, and age, word recognition, listening comprehension, and the interaction between word recognition and listening comprehension as predictors. The complete model accounted for 41.9% of the variance. Changes in $\underline{R}^2$ at each step of the two analyses are presented in Table 4.9. In both analyses, word recognition and listening comprehension contributed significantly to unaided reading comprehension, independent of age and each other. The interaction variable did not contribute to unaided reading comprehension beyond the independent variables.

Table 4.9. Hierarchical multiple regression analyses conducted in Experiment 10, with unaided reading comprehension as the dependent variable.

Step	Added variable	Change in R^2	Significance of change in R^2
1	Age	.01	non-significant
2	Listening comprehension	.14	$p<.001$
3	Word recognition	.26	$p<.001$
2	Word recognition	.34	$p<.001$
3	Listening comprehension	.06	$p<.01$
4	Word recognition x listening comprehension	.00	non-significant

Having demonstrated that word recognition and listening comprehension were independent contributors to unaided reading comprehension ability, the final stage of analyses examined the independence of the reading accuracy and reading comprehension measures provided by the Neale Analysis of Reading Ability. Two non-linear hierarchical multiple regression analyses were conducted as above. Unaided reading comprehension was the dependent variable, and the predictors were age, reading accuracy score and reading comprehension score obtained on the Neale Analysis of Reading Ability, and the interaction between reading accuracy and reading comprehension scores. The complete model accounted for 39.4% of the variance. Changes in R^2 at each step of the two analyses are provided in Table 4.10. Reading accuracy contributed significantly to unaided reading comprehension, independent of age and reading comprehension. The contribution of reading comprehension performance on the Neale Analysis of Reading Ability to unaided reading comprehension independent of age and reading accuracy was not significant. The interaction variable did not contribute to unaided reading comprehension beyond the independent variables.

Table 4.10. Hierarchical multiple regression analyses conducted in Experiment 10, with unaided reading comprehension as the dependent variable.

Step	Added variable	Change in R^2	Significance of change in R^2
1	Age	.01	non-significant
2	Reading comprehension	.20	$p<.001$
3	Reading accuracy	.17	$p<.001$
2	Reading accuracy	.36	$p<.001$
3	Reading comprehension	.02	non-significant
4	Reading accuracy x reading comprehension	.00	non-significant

In summary, some children who exhibited low performance on the reading accuracy or reading comprehension measures provided by the Neale Analysis of Reading Ability did not perform poorly on alternative tests of reading or comprehension ability. It was also found that performance on the Neale Analysis of Reading Ability could not reliably predict performance on the alternative measures of reading and comprehension ability. Furthermore, whilst reading and comprehension abilities were demonstrated to be independent contributors to unaided reading comprehension, comprehension score obtained on the Neale Analysis of Reading Ability did not contribute to unaided reading comprehension independent of reading accuracy score obtained on the Neale Analysis of Reading Ability.

4.4.3 Discussion

Experiment 10 was conducted to investigate the validity of using the Neale Analysis of Reading Ability to identify groups of different reading profiles. Comparisons were conducted between patterns of performance across the reading accuracy and reading comprehension measures of the Neale Analysis of Reading Ability and across two alternative measures of reading and comprehension ability, which did not feature aspects of concern present in the Neale Analysis of Reading Ability.

In the first stage of analysis, groups of skilled readers, less-skilled comprehenders and less-skilled decoders were selected according to their performance on the Neale Analysis of Reading Ability. The performances of each group across the two alternative measures were then inspected. The skilled reading group tended to score above average on the alternative measures of both reading and comprehension ability. This finding suggests that the Neale Analysis of Reading Ability provided valid identification of children who were skilled in both reading and comprehension. The groups of less-skilled comprehenders scored above average on the alternative measure of reading ability, but did not reliably score below average on the alternative measure of comprehension ability. This finding suggests that the comprehension ability of the latter children was underestimated by the Neale Analysis of Reading Ability. It would, therefore, appear that if the Neale Analysis of Reading Ability is used to identify groups of less-skilled comprehenders – children with poor text comprehension despite good reading accuracy – the groups may contain children whose comprehension is not, in fact, impaired.

Similarly, the groups of less-skilled decoders – selected according to performance on the Neale Analysis of Reading Ability – exhibited equivalent tendency to score above average on the alternative comprehension measure and below average on the alternative reading measure, and to score above average on both alternative measures. This finding suggests that the measure of reading ability provided by the Neale Analysis of Reading Ability can underestimate reading ability. If the Neale Analysis of Reading Ability is used to identify groups of less-skilled decoders, the groups may contain children whose reading ability is not impaired.

The next stage of analysis investigated whether the Neale Analysis of Reading Ability could reliably identify children who exhibited certain profiles across the two alternative measures of reading and comprehension ability. The Neale Analysis of Reading Ability provided good classification of children who scored highly on the alternative measures of reading and comprehension, but only moderate classification of children who scored highly on the alternative

reading measure and low on the alternative comprehension measure. This would appear to reflect the unreliable assessment of comprehension ability provided by the Neale Analysis of Reading Ability, as discussed above. Furthermore, the Neale Analysis of Reading Ability provided very poor classification of children who scored well on the alternative measure of comprehension and low on the alternative reading measure, and this would appear to reflect the unreliable assessment of reading ability provided by the Neale Analysis of Reading Ability, as discussed above. In short, the Neale Analysis of Reading Ability appears able to identify skilled reading and skilled comprehension, but does not exhibit reliable identification of poor reading or poor comprehension ability.

The final phase of Experiment 10 investigated whether reading ability and comprehension ability combine to improve unaided reading comprehension, or whether their contribution to reading comprehension is strictly additive. It was found that reading and comprehension abilities were independent contributors to reading comprehension ability. This finding validates the selection of groups of children with distinct profiles across reading ability and comprehension ability, for the experimental investigation into what underlies these skills. However, it was also found that comprehension score obtained on the Neale Analysis of Reading Ability did not contribute to reading comprehension independent of reading accuracy score obtained on the Neale Analysis of Reading Ability. This finding demonstrates that the Neale Analysis of Reading Ability does not provide independent assessments of reading and comprehension abilities. It follows, therefore, that the Neale Analysis of Reading Ability might not provide valid identification of groups with distinct profiles across reading and comprehension.

In short, Experiment 10 found that reading ability and comprehension ability contribute independently to reading comprehension. This validates the selection of groups of specific profiles across reading and comprehension. However, it was found that the Neale Analysis of Reading Ability did not provide independent assessments of reading and comprehension ability, and did not

appear able to reliably identify poor performance. These findings cast doubt over the validity of the Neale Analysis of Reading Ability at selecting such groups.

4.5 General Discussion

The studies reported in this chapter assessed the validity of the Neale Analysis of Reading Ability (Neale, 1997) in identification of children with distinct profiles across reading and comprehension abilities. This practice assumes that reading and comprehension abilities are independent skills, so that children can be good at one and poor at the other. It was demonstrated in Experiment 10 that reading and comprehension are indeed independent abilities, validating the investigation of distinct profiles. However, it was also demonstrated in Experiment 10 that the comprehension assessment provided by the Neale Analysis of Reading Ability is not independent of the reading accuracy assessment provided by this test. Since the Neale Analysis of Reading Ability does not provide independent measures of reading and comprehension abilities, it does not appear to enable valid identification of children with distinct profiles across these components of reading comprehension.

This chapter also addressed concerns about the two separate measures of the Neale Analysis of Reading Ability. Firstly, evidence suggested that the Neale Analysis of Reading Ability was able to identify good reading ability. Experiments 8 and 9 found good correspondence between performance on the reading accuracy measure of the Neale Analysis of Reading Ability and an alternative test of reading ability, in the general population and in children who scored well on the reading accuracy measure. However, Experiments 8 and 10 both found that not all children who scored poorly on the reading accuracy measure of the Neale Analysis of Reading Ability continued to score poorly on the alternative measure of reading ability. The primary difference between the two measures of reading ability provides a possible account for this inconsistent performance. Although the assessment provided by the Neale Analysis of Reading Ability is referred to that of reading ability, it actually assesses ability to produce the sounds of written words - decoding . The alternative measure, on

the other hand, made no decoding demands. Decoding requires vocalisation of text in addition to reading, and some children may have difficulty with the vocalisation of text despite adequate reading ability. Children with physical difficulties formulating sounds will be at an obvious disadvantage, as will careless decoders and children with low confidence. Such children might be expected to achieve a score on the reading accuracy measure of the Neale Analysis of Reading Ability that underestimates their ability to recognise written words, and the findings from Experiments 8 and 10 provided some evidence that this occurs.

It appears that some children score can poorly on the reading accuracy measure of the Neale Analysis of Reading Ability for reasons other than poor reading ability. Furthermore, it was demonstrated in Experiment 10 that the Neale Analysis of Reading Ability provided low classification of children who exhibited poor word recognition. This finding reflects the low certainty that poor performance on the reading accuracy measure of the Neale Analysis of Reading Ability indicates knowledge of written words. It follows, therefore, that if the Neale Analysis of Reading Ability is used to identify groups of less-skilled decoders, such groups might derive more from written words than their original assessment suggested.

This chapter also addressed concerns about the comprehension measure provided by the Neale Analysis of Reading Ability. Experiment 9 demonstrated good correspondence between performance on the reading comprehension measure of the Neale Analysis of Reading Ability and listening comprehension, in the general population. However, it was found in Experiments 9 and 10 that the Neale Analysis of Reading Ability was unable to predict listening comprehension ability, and provided poor classification of children who scored poorly on listening comprehension. These findings are consistent with the proposal that the reading comprehension assessment of the Neale Analysis of Reading Ability makes more demands than does the listening comprehension assessment, and, therefore, that listening comprehension offers a more accurate assessment of comprehension ability.

This proposal is further supported by the finding that many children in Experiments 9 and 10 scored well on the listening comprehension measure despite having scored very poorly in the comprehension measure of the Neale Analysis of Reading Ability. There are several possible accounts for this pattern of performance. Firstly, in addition to sufficient text comprehension, the comprehension measure of the Neale Analysis of Reading Ability requires oral reading to be efficient enough that forming the sounds of words does not detract attention from text content. The comprehension measure also requires oral reading to be sufficient accurate, permitting children to attempt a reasonable number of questions before reaching the accuracy cut-off. It was found in Experiment 9 that a number of the children who scored poorly on the comprehension measure of the Neale Analysis of Reading Ability but well on listening comprehension assessment exhibited low reading accuracy. In addition, the finding in Experiment 10 that comprehension score was somewhat dependent on reading accuracy score in the Neale Analysis of Reading Ability implies that comprehension score might indeed be limited by reading accuracy. Secondly, the comprehension measure of the Neale Analysis of Reading Ability required children to answer open-ended questions. These questions demand skills that are additional to comprehension, including sufficient meta-comprehension, expressive speech, and even confidence. It was demonstrated in Experiment 6 that some children do score poorly on the open-ended questions without exhibiting poor comprehension when assessed by forced-choice questions. In short, findings from Experiments 9 and 10 suggest that the reading comprehension measure of the Neale Analysis of Reading Ability makes demands that are additional to comprehension – requiring the formation of expressive answers and relying on reading accuracy – and that these demands risk underestimating individual's comprehension ability.

To summarise, Chapter 4 presented evidence that children can score poorly on the reading accuracy or reading comprehension measures of the Neale Analysis of Reading Ability for reasons other than poor reading or poor comprehension. This implies that if the Neale Analysis of Reading Ability is used to select groups of children with poor reading or comprehension ability, the samples

might be polluted by children who are not, in fact, impaired on these skills, but who, instead, have some difficulty with the extra demands made by the Neale Analysis of Reading Ability. Furthermore, it was found that reading and comprehension are independent skills, but that the comprehension assessment of the Neale Analysis of Reading Ability is not independent of the reading assessment provided by the Neale Analysis of Reading Ability. The Neale Analysis of Reading Ability does not appear to offer an accurate reflection of separate reading and comprehension abilities, and therefore appears inappropriate as an analytic tool of assessment.

CHAPTER 5

GENERAL DISCUSSION

5.1 Poor comprehension, integration and memory

In Experiments 1 to 5, a sentence recognition test (c.f., Oakhill, 1982; Oakhill et al., 1986) was used to explore whether poor integration of information or poor memory lie at the root of poor comprehension ability. The experiments found no experimental differences between skilled and less-skilled comprehenders. This has three possible explanations. The first explanation is that poor integration and poor memory do not underlie poor comprehension. The second explanation is that the sentence recognition test was not sensitive to differences between the two groups, and the third is that the groups who participated in these experiments were not representative of children with genuine good or low comprehension ability. These latter two possibilities shall now each be discussed in turn.

5.1.1 The sentence recognition test

In this section it is discussed whether the absence of group differences on the sentence recognition test in Experiments 1 to 5 may reflect a weakness in that test, in particular that the sentence recognition test was not sensitive to cognitive differences between children of good and low comprehension ability. The sentence recognition test involved processing of simple three-line stories, and it might therefore be the case that children who exhibited poor comprehension during screening on the Neale Analysis of Reading Ability found the sentence recognition test stories within their ability. Experiment 6 indeed found that less-skilled comprehenders exhibited comprehension at the level of the skilled group over the first two stories of the Neale Analysis of Reading Ability. Since these stories were not even as short and simple as the sentence recognition test stories, it does indeed appear that the sentence recognition test was too easy to be sensitive to deficiencies in the less-skilled group. The null effects in Experiments 1 to 5 may, therefore, reflect insensitivity of the experimental test.

5.1.2 The Neale Analysis of Reading Ability comprehension assessment

Another possibility is that the lack of group differences in Experiments 1 to 5 reflects poor group selection. In particular, the inclusion in the less-skilled comprehension group of children who scored poorly on the Neale Analysis of Reading Ability (Neale, 1997) for reasons other than poor comprehension would reduce the difference between the groups on a test (i.e., the sentence recognition test) that was designed to assess skills underlying comprehension. In two ways the comprehension measure of the Neale Analysis of Reading Ability makes demands that are additional to comprehension. There is a risk, therefore, that some children score poorly due to these additional demands rather than due to poor comprehension per se.

The first way in which the comprehension measure makes additional demands lies in the type of questioning used to assess comprehension. Children are required to provide descriptive, explicative answers to open-ended questions. It was argued in Chapter 3 that these kinds of questions require meta-comprehension, expressive speech, and even sufficient confidence, and that a child with some difficulty formulating answers to such questions would be impaired beyond his/her comprehension ability. It is, therefore, possible that the less-skilled comprehension groups in Experiments 1 to 5 included some children who scored poorly on the Neale Analysis of Reading Ability due to difficulties with open-ended questions, rather than comprehension per se. In Experiment 6 this possibility was investigated, administering a parallel version of the Neale Analysis of Reading Ability in which the question demands had been reduced to a forced choice veracity judgment. It was found that groups of skilled and less-skilled comprehenders, identified in the same manner as in Experiments 1 to 5, no longer differed in their comprehension score when the question demands were reduced in this manner. In addition, it was found in Experiments 9 and 10 that children who scored poorly on the traditional version of the Neale Analysis of Reading Ability comprehension measure did not always score poorly on a similar listening comprehension test, in which the questions were of the true/false type. These findings strongly suggest that the Neale Analysis of Reading Ability may identify as less-skilled comprehenders

some children whose comprehension is not, in fact, impaired, but who instead have some difficulty formulating answers about their comprehension.

The second way in which the comprehension measure of the Neale Analysis of Reading Ability may underestimate comprehension is that it makes demands on reading. Firstly, some children may obtain low comprehension score on the Neale Analysis of Reading Ability because their resources were largely directed to reading. While the groups in Experiments 1 to 5 were matched for reading accuracy score on the Neale Analysis of Reading Ability, it is possible that some children included in less-skilled comprehension groups obtained good reading accuracy scores at the cost of their comprehension. Experience administering the Neale Analysis of Reading Ability has revealed that children differ in their attention to reading accuracy – some racing ahead regardless of accuracy, others slowly decoding every word. Children who adopt the second strategy may score well on reading accuracy through effort, while that effort detracted from their attention to comprehending the text. Secondly, children are only permitted to attempt comprehension questions when the number of reading errors they make is below a certain cut-off value. Children with low reading accuracy are therefore restricted in the comprehension score that they may obtain, regardless of their ability to understand stories.

Several findings from this thesis are relevant to the demands placed by reading for comprehension in the Neale Analysis of Reading Ability. On the one hand, findings from Experiments 6 and 7 suggested that skilled and less-skilled comprehenders – supposedly matched for reading accuracy – exhibited general comprehension ability between reading aloud and reading silently. This suggests that those groups – selected in the same manner as those in Experiments 1 to 5 – were not affected by the additional demands of reading aloud in the Neale Analysis of Reading Ability. However, these findings may reflect continued effect of reading demand on the comprehension performance of the less-skilled comprehension group, regardless of whether children need to vocalise their reading; The less-skilled comprehension group may still be directing attention to reading at the cost of their comprehension. Other findings

support this interpretation. The findings of Experiment 10 demonstrated across a large sample that the Neale Analysis of Reading Ability comprehension measure was dependent on reading accuracy. Furthermore, it was demonstrated in Experiments 9 and 10 that children could exhibit poor comprehension when reading the Neale Analysis of Reading Ability and good comprehension in a similar test which removed reading demands through its administration as a listening test.

In short, evidence has suggested that children who scored poorly on the comprehension assessment of the Neale Analysis of Reading Ability may have done so because of additional reading demands. Experiments 1 to 5 were conducted to investigate comprehension ability, and it is undesirable, therefore, that the assessment of comprehension ability was confounded with reading ability. Whilst it was assumed that matching the groups for reading accuracy would control for reading ability, it was later realised that some children may have obtained good reading accuracy score at the cost of their comprehension, and that, therefore, the less-skilled comprehension group may have included some children whose comprehension deficit was overestimated. This suspicion was supported by findings in Experiment 10 that even when groups were selected on the basis of both comprehension and reading accuracy, some children in the less-skilled comprehension group scored well when administered a comprehension test that removed reading demands.

To summarise, the lack of group differences in Experiments 1 to 5 may have resulted from poor group selection. Evidence has suggested that children can score poorly on the comprehension assessment of the Neale Analysis of Reading Ability for reasons other than poor comprehension ability. In particular, the Neale Analysis of Reading Ability confounds comprehension with reading ability, and requires sufficient expressive speech to answer the questions. The less-skilled comprehension groups of Experiments 1 to 5 may, therefore, have included children whose comprehension was not, in fact, impaired, but who, instead, had some difficulty with reading or with

formulating descriptive answers. This would have reduced group differences on the sentence recognition test.

5.1.3 Poor comprehension, integration and memory - summary

Section 5.1 reviewed possible explanations for consistent failure in Experiments 1 to 5 to find expected differences between skilled and less-skilled comprehenders on a sentence recognition test, that was designed to tap integration of information and memory. The first possible account for this pattern of results was that poor comprehension does not arise from poor integration or poor memory. The second account might be that the sentence recognition test was not sensitive to cognitive deficits in less-skilled comprehenders, because the materials were very simple, and within the ability of the less-skilled group. Findings have suggested that this is the case. The third account is one of poor group selection. Evidence has strongly suggested that the deficit of the less-skilled group may have been overestimated due to the extra demands made by the Neale Analysis of Reading Ability beyond comprehension. In particular, evidence suggested that children can score poorly on the comprehension measure of the Neale Analysis of Reading Ability because of poor expressive speech or poor reading ability, rather than poor comprehension ability per se.

5.2 The Neale Analysis of Reading Ability reading assessment

Having raised concerns about the comprehension assessment of the Neale Analysis of Reading Ability (Neale, 1997), the validity of the reading ability assessment of this test was then examined. In Chapter 4 it was discussed that while the Neale Analysis of Reading Ability offers an assessment of reading ability – ability to recognise and know the meanings of written words – it actually assesses decoding – ability to access the sounds of written words. The use of the Neale Analysis of Reading Ability, therefore, assumes that reading and decoding abilities are the same. However, some children may 'bark' at text – saying the sounds of words without processing their meanings – while other children may recognise the meanings of written words but have some difficulty accessing their sound.

While these cases are hypothetical, work conducted in this thesis yielded evidence that abilities to recognise written words and to decode do not always concur. A number of children were found in Experiments 8 and 9 who scored poorly on the Neale Analysis of Reading Ability decoding measure despite good recognition of written words. This evidence suggests that some children may obtain low scores on the reading ability measure of the Neale Analysis of Reading Ability through additional vocalising demands, rather than poor reading ability per se.

This thesis is not primarily concerned with reading, but with processes underlying the construction of comprehension from series of words. In line with existing research, the Neale Analysis of Reading Ability was adopted to provide an assessment of comprehension ability while controlling for reading ability. This practice assumed that children who obtained good reading scores on the Neale Analysis of Reading Ability were children who recognised the meanings of individual words. The less-skilled comprehension groups in Experiments 1 to 5 were, therefore, assumed to offer a sample of children who had some deficit in the chain of processes from meaning at the most basic level to meaning of text or discourse. However, the evidence discussed in this section suggests that this assumption may have been premature – there is no evidence that the less-skilled comprehension groups had abstracted meaning even at the level of single words. Such information may have been obtained by a vocabulary test, such as the British Picture Vocabulary Scales (Dunn et al., 1997), which might be administered as a reading vocabulary test, or by a word discrimination task, such as the Word Decision Test (Baddeley et al., in press).

In short, evidence has suggested that the distinction between word recognition and decoding is relevant to assessment of reading ability. Since the Neale Analysis of Reading Ability was used throughout the work presented in this thesis, it appears that the comprehension groups in Experiments 1 to 5 may not have been matched on understanding at the single word level, as intended. It was proposed that alternative assessments, such as vocabulary or word discrimination tests, should be used in future work.

5.3 Comprehension ability as opposed to test performance

Investigations into the deficits underlying poor ability to comprehend primarily require an accurate assessment of comprehension ability. This section reviews the evidence from this thesis that such assessments can depend on task demands, and then discusses the impact that this has on research into comprehension ability.

5.3.1 Reading condition

The first manipulation of comprehension performance was conducted in Chapter 3, between reading aloud and corrected, as in the Neale Analysis of Reading Ability, and reading silently and independently. In Experiments 6 and 7 it was demonstrated that skilled readers and less-skilled comprehenders exhibited equivalent comprehension across reading condition, but that the comprehension of less-skilled decoders was impaired by reading silently. It appears that the children with good reading ability – the skilled readers and less-skilled comprehenders – were unaffected by reading condition, presumably because their reading was sufficiently accurate to provide them with the necessary information. The children with poor reading ability, on the other hand, were presumed to comprehend poorly when reading silently and unaided because their poor reading accuracy meant that they were not able to provide themselves with the necessary word information from which to build comprehension.

It appears, therefore, that an individual with poor reading is likely to exhibit different comprehension between tests that require independent silent reading and those in which the individual reads aloud with correction and assistance. It is important to realise that such tests are not assessing children's general comprehension ability, but their ability to comprehend what they can read. Such tests do not, therefore, offer an assessment of pure comprehension.

5.3.2 Test versus independent class work

In Experiment 6 it was investigated how comprehension performance on the Neale Analysis of Reading Ability related to the comprehension exhibited when

working independently from text in a classroom situation. Skilled readers, who scored well on both the reading and comprehension measures of the Neale Analysis of Reading Ability, were used as controls, and the performance of children with specific profiles on the Neale Analysis of Reading Ability was compared to the skilled readers on a classroom comprehension assessment.

It was found that the Neale Analysis of Reading Ability can overestimate the classroom comprehension of children with poor reading accuracy, because the Neale Analysis of Reading Ability provides assistance that is not available during independent class work – see Section 5.3.1. It was also demonstrated that some children with good decoding accuracy can score more highly in the classroom than predicted by the Neale Analysis of Reading Ability. One reason for this may be that the classroom comprehension assessment used *true/false* questions, and it was also demonstrated in Experiment 6 that some children appear to obtain low scores on the Neale Analysis of Reading Ability due to the additional demands that those questions make (see Section 5.1.2). An additional reason for this pattern of performance may be the change in conditions. Children completed the classroom assessment at their own pace, and this allowed children to read back over poorly comprehended sections, filling gaps in their comprehension or amending mistakes. Furthermore, children were no longer required to work under the watchful eye of the examiner, which some children may have found very intimidating, impairing their performance.

These findings again demonstrate that comprehension performance can differ between test condition. Furthermore, these findings suggest that performance on the Neale Analysis of Reading Ability cannot always provide an accurate reflection of comprehension in real life situations.

5.3.3 The Neale Analysis of Reading Ability versus alternative tests

In Chapter 4 it was investigated how the performance on the assessments of reading and comprehension provided by the Neale Analysis of Reading Ability relate to alternative measures of reading and comprehension. It was demonstrated in Experiments 8 and 10 that some children can score poorly on the reading assessment of the Neale Analysis of Reading Ability and obtain high score on the Word Decision Test (Baddeley et al., in press) – a test that makes no decoding and no comprehension demands. In addition, it was demonstrated in Experiments 9 and 10 that performance on the comprehension assessment of the Neale Analysis of Reading Ability does not always correspond to performance on an alternative comprehension test, that made no reading demands and reduced the question demands. These findings further support the claim that performance on a test is affected by the demands made by that test that are additional to the target skills. In particular, the Neale Analysis of Reading Ability places several demands that are additional to reading and comprehension, and the assessments offered by this test do not, therefore, provide a scientifically accurate measurement of reading and comprehension ability.

5.3.4 Comprehension ability versus performance - summary

This thesis has provided evidence that comprehension performance can be affected by test demands. This feature is problematic for research into comprehension ability, as it is unclear exactly which test would provide the most accurate assessment of that ability. In particular, one might wish to avoid assessments which risk low performance due to difficulties with tasks not directly related to comprehension, such as reading or expressive speech. Whilst this thesis aimed to investigate comprehension ability, it had in fact been unable to accurately identify comprehension ability. A new test would be needed, that minimises the additional demands, and thereby approaches more closely the ability that underlies comprehension performance on other tests.

5.4 Future comprehension test

The Neale Analysis of Reading Ability (Neale, 1997) is a very useful tool of assessment in education. As a standardised and widely used test, it provides a comparison between children's ability to explain their understanding of texts that they can read. This set of skills closely resembles successful reading for comprehension at school. This is particularly true of interactive situations, where children can ask for assistance, where questions can be rephrased when necessary, and where children's expressive speech is an important aspect of communication in the classroom. The Neale Analysis of Reading Ability, therefore, provides an important and valuable assessment of a child's successful reading for comprehension in an interactive educational situation. This thesis does not criticise the use of the Neale Analysis of Reading Ability in education.

However, the comprehension assessment provided by the Neale Analysis of Reading Ability does have some problems for scientific use, since it involves too many factors. As discussed in Section 5.3, investigations into comprehension ability should not depend on other skills, but evidence has shown (see Section 5.1.2) that performance on the comprehension assessment of the Neale Analysis of Reading Ability can be affected by other cognitive difficulties. Some of the findings of this thesis can suggest a test format that might provide an assessment that is closer to actual comprehension ability than that provided by the Neale Analysis of Reading Ability. Given the findings of Experiment 6 that open-ended questions require more than comprehension, and the findings of Experiments 6, 7, 9 and 10 that reading or decoding for comprehension require more than comprehension, an ideal future test of comprehension would reduce question demands and remove all reading demands. The evidence therefore suggests that a better assessment of comprehension would be provided by a listening test with *true/false* questions, such as the test developed in Section 4.1

5.5 Future directions

The question of whether poor integration or poor memory underlie less-skilled comprehension is still important and interesting, and merits investigation.

Further work leading from this thesis would be, therefore, to conduct the comparisons of Experiments 1 to 5 again, improving both group selection and experimental materials. The findings presented in this thesis would suggest that the materials should be of complexity no easier than those of Level 3 in the Neale Analysis of Reading Ability. Furthermore, evidence suggests that the alternative test proposed in Section 5.4 should be developed and applied to group selection in the investigation of skills underlying comprehension ability.

In addition to using the listening comprehension test described in Section 5.4 to select groups in further work, it might be prudent to also control for vocabulary (as in Experiment 5), and for comprehension of individual sentences. This strategy would provide a more accurate identification of groups of less-skilled comprehenders whose deficit is limited to some aspect of the processes involved in comprehension between sentences and discourse. By selecting groups in this manner, research questions could be targeted more specifically at text or discourse comprehension.

Other interesting research could investigate the contributions of a number of skills to comprehension ability within large samples. One avenue of work would be to explore the relative contributions of various skills to comprehension ability, indicating their importance to skilled comprehension, and providing more information about what is involved in comprehension. Furthermore, including samples of different ages would indicate the importance of different skills as children get older and their comprehension improves. It might be found that skilled comprehension does not arise from one specific pattern of weight across the various component skills, but rather that a number of successful profiles exist.

Finally, it would be interesting and important to investigate whether less-skilled comprehension has a number of different causes. It is possible that of the children with poor comprehension ability, not all have the same cognitive deficits underlying their comprehension impairment. It would be interesting to investigate the relative occurrences of the different underlying deficits, and it

would be important to enable identification of specific causes of poor comprehension, in order to then devise and develop remedial strategies that are tailored to different needs.

REFERENCES

Adams, A.-M., Bourke, L., & Willis, C. (1999). Working memory and spoken language comprehension in young children. International Journal of Psychology, 34, 364-373.

Anderson, R. C., & Ortony, A. (1975). On putting apples into bottles - a problem of polysemy. Cognitive Psychology, 7, 167-180.

Baddeley, A. (1986). Working memory. Oxford: Oxford University Press.

Baddeley, A. (1996). The concept of working memory. In S. E. Gathercole (Ed.), Models of short-term memory (pp. 1-22). Sussex: Erlbaum.

Baddeley, A. D., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. Quarterly Journal of Experimental Psychology, 33, 439-454.

Baddeley, A. D., Gathercole, S. E., & Spooner, A. L. R. (in press). The Reading Decision Test. Brighton: Psychology Press.

Bartlett, F. C. (1932). Remembering: A study in experimental and social psychology. Cambridge: Cambridge University Press.

Bishop, D. D. V. (1989). Test for reception of grammar (TROG), Available from the MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. Cognitive Psychology, 3, 193-209.

Bransford, J. D., & Franks, J. J. (1971). The abstraction of linguistic ideas. Cognitive Psychology, 2, 331-350.

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. Reading and Writing, 11, 489-503.

Cain, K., Oakhill, J., & Sharp, N. (1998). Working memory, reading comprehension and the ability to learn new vocabulary items from context. Paper presented at the BPS Developmental Section, Lancaster.

Carr, T. H., Brown, T. L., Vavrus, L. G., & Evans, M. A. (1990). Cognitive skill maps and cognitive skill profiles: Componential analysis of individual differences in children's reading efficiency. In T. H. Carr &

B. A. Levy (Eds.), Reading and Its Development: Component Skills Approaches (pp. 1-55). Academic Press.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behaviour, 19, 450-466.

Daneman, M., & Merickle, P. M. (1996). Working memory and language comprehension: A meta-analysis. Psychonomic Bulletin and Review, 3, 422-433.

De Beni, R., Pallandino, P., Pazzaglia, F., & Cornoldi, C. (1998). Increases in intrusion errors and working memory deficit of poor comprehenders. Quarterly Journal of Experimental Psychology, 51, 305-320.

Dixon, P., LeFevre, J.-A., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. Journal of Educational Psychology, 80, 465-472.

Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). The British Picture Vocabulary Scale, second edition. Windsor: National Foundation for Educational Research.

Estes, W. K. (1973). Phomenic coding and rehearsal in short-term memory. Journal of Verbal Learning and Verbal Behavior, 12, 360-372.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The children's test of nonword repetition: A test of phonological working memory. Memory, 2, 103-127.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. Journal of Experimental Psychology: Learning, Memory and Cognition, 16, 430-445.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. Psychological Review, 101, 371-395.

Hannon, B., & Daneman, M. (1998). Facilitating knowledge-based inferences in less-skilled readers. Contemporary Educational Psychology, 23, 149-172.

Johnson-Laird, P. N. (1985). Mental models. Second edition. Cambridge: University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. Psychological Review, 95, 163-182.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. Psychological Review, 85, 363-394.

Leather, C. V., & Henry, L. A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. Journal of Experimental Child Psychology, 58, 88-111.

Levy, B. A. (1971). The role of articulation in auditory and visual short-term memory. Journal of Verbal Learning and Verbal Behavior, 10, 123-132.

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 20, 1456-1470.

McKoon, G., & Ratcliff, R. (1992). Inference during reading. Psychological Review, 99, 440-466.

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during narrative comprehension. Journal of Memory and Language, 28, 292-312.

Murray, D. J. (1967). The role of speech in short-term memory. Canadian Journal of Psychology, 21, 263-276.

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. Journal of Experimental Child Psychology, 73, 139-158.

Nation, K., Marshall, C. M., & Snowling, M. J. (2001). Phonological and semantic contributions to children's picture naming skill: Evidence from children with developmental reading disorders. Language and Cognitive Processes, 16, 241-259.

Nation, K., & Snowling, M. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. British Journal of Educational Psychology, 67, 359-370.

Nation, K., & Snowling, M. (1998a). Semantic processing and the development of word-recognition skills: Evidence from children with reading

comprehension difficulties. Journal of Memory and Language, 39, 85-101.

Nation, K., & Snowling, M. J. (1998b). Individual differences in contextual facilitation: Evidence from dyslexia and poor reading comprehension. Child Development, 69, 996-1011.

Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. Applied Psycholinguistics, 21, 229-241.

Neale, M. D. (1958). Neale Analysis of Reading Ability Manual. London: Macmillan.

Neale, M. D. (1997). Neale Analysis of Reading Ability - Revised. Windsor: NFER-NELSON.

Neale, M. D., Christophers, U., & Whetton, C. (1989). Neale Analysis of Reading Ability - Revised British Edition. Windsor: NFER - NELSON.

Oakhill, J. (1982). Constructive processes in skilled and less skilled comprehenders' memory for sentences. British Journal of Psychology, 73, 13-20.

Oakhill, J. (1983). Instantiation in skilled and less skilled comprehenders. Quarterly Journal of Experimental Psychology, 35, 441-450.

Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. British Journal of Educational Psychology, 54, 31-39.

Oakhill, J., Yuill, N., & Parkin, A. (1986). On the nature of the difference between skilled and less-skilled comprehenders. Journal of Research in Reading, 9, 80-91.

Paris, S. G., & Carter, A. Y. (1973). Semantic and constructive aspects of sentence memory in children. Developmental Psychology, 9, 109-113.

Pickering, S., & Gathercole, S. (2001). Working Memory Test Battery for Children. London: Psychological Corporation.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 2, 437-442.

Seigneuric, A., Ehrich, M.-F., Oakhill, J. V., & Yuill, N. M. (2000). Working memory resources and children's reading comprehension. Reading and Writing: An Interdisciplinary Journal, 13, 81-103.

Stothard, S. E., & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. Reading and Writing, 4, 245-256.

Stothard, S. E., & Hulme, C. (1995). A comparison of phonological skills in children with reading comprehension difficulties and children with decoding difficulties. *Journal of Child Psychological Psychiatry, 36*(3), 399-408.

Yuill, N., Oakhill, J., & Parkin, A. (1989). Working memory, comprehension ability and the resolution of text anomaly. British Journal of Psychology, 80, 351-361.

APPENDICES

Appendix A. Test materials used in Experiment 2

Story set A

Stories.

1 The man kicked the ball.
The ball was on the grass.
The man fell into the water.

2 The lady bought a present.
The present was a book.
The lady looked for a game.

3 The girl stood next to the boy.
The boy was in the garden.
The girl looked at the clouds.

4 The pilot flew the plane.
The plane was going to Australia.
The pilot wanted to be on holiday.

5 The sheep bumped into the cow.
The cow was in a field.
The sheep slipped on the mud.

6 The chef cooked the dinner.
The dinner was sausages.
The chef looked for some chips.

7 The policeman caught the thief.
The thief was a tall man.
The policeman ran very fast.

8 The pepper fell onto the butter.
The butter was on the table.
The pepper spilled everywhere.

Recognition sentences.

1.1 The man kicked the ball.

1.2 The ball was on the grass.

1.3 The man was on the grass.

1.4 The ball fell into the water.

2.1 The lady looked for a game.

2.2 The present was a book.

2.3 The lady bought a book.

2.4 The present was a game.

3.1 The girl stood next to the boy.

3.2 The boy was in the garden.

3.3 The girl was in the garden.

3.4 The boy looked at the clouds.

4.1 The pilot flew the plane.

4.2 The plane was going to Australia.

4.3 The pilot was going to Australia.

4.4 The pilot was on holiday.

5.1 The sheep bumped into the cow.

5.2 The cow was in a field.

5.3 The sheep was in a field.

5.4 The cow slipped on the mud.

6.1 The chef looked for some chips.

6.2 The dinner was sausages.

6.3 The chef cooked the sausages.

6.4 The dinner was chips.

7.1 The policeman ran very fast.

7.2 The thief was a tall man.

7.3 The policeman caught the tall man.

7.4 The policeman was a tall man.

8.1 The pepper spilled everywhere.

8.2 The butter was on the table.

8.3 The pepper was on the table.

8.4 The butter fell onto the pepper.

Story set B

Stories.

1 The dog chased the cat.

The cat was in a field.

The dog saw a rabbit.

2 The waiter spilled the drink.

The drink was milk.

The waiter got some water.

3 The teacher reached for a pen.

The pen was in the drawer.

The teacher found a ruler.

4 The boy got breakfast.

The breakfast was some toast.

The boy wanted Cornflakes.

5 The girl watched the film.

The film was The Lion King.

The girl wanted the Jungle Book.

6 The nurse helped the doctor.

The doctor was in the hospital.

The nurse got the medicine.

7 The train left the station.

The station was at Birmingham.

The train went to London.

8 The man gave the umbrella to the lady.

The lady was in the rain.

The man got soaked.

Recognition sentences.

1.1 The dog saw a rabbit.

1.2 The cat was in a field.

1.3 The dog was in a field.

1.4 The cat saw a rabbit.

2.1 The waiter spilled the drink.

2.2 The drink was milk.

2.3 The waiter spilled the milk.

2.4 The drink was water.

3.1 The teacher found a ruler.

3.2 The pen was in the drawer.

3.3 The ruler was in the drawer.

3.4 The teacher reached for a ruler.

4.1 The boy got breakfast.

4.2 The breakfast was some toast.

4.3 The boy got some toast.

4.4 The breakfast was Cornflakes.

5.1 The girl wanted the Jungle Book.

5.2 The film was The Lion King.

5.3 The girl watched The Lion King.

5.4 The film was the Jungle Book.

6.1 The nurse helped the doctor.

6.2 The doctor was in the hospital.

6.3 The nurse was in the hospital.

6.4 The doctor got the medicine.

7.1 The train left the station.

7.2 The station was at Birmingham.

7.3 The train left Birmingham.

7.4 The station was at London.

8.1 The man got soaked.

8.2 The lady was in the rain.

8.3 The man was in the rain.

8.4 The lady gave the umbrella to the man.

Appendix B. Test materials used in Experiments 3, 4 and 5.

Story set A

Stories.

1 The dog wanted the dinner.

The dinner was steak.

The dog got some biscuits.

2 The car drove over the hat.

The hat was in the road.

The car parked in the garage.

3 The cow stood next to the sheep.

The sheep was in the farm.

The cow ate some grass.

4 The bike crashed into the van.

The van was near the playground.

The bike skidded on the gravel.

5 The cat chased the crow.

The crow was in a field.

The cat hid in a bush.

6 The TV fell onto the video.

The video was in the living room.

The TV smashed to pieces.

7 The rabbit ate the food.

The food was carrots.

The rabbit looked for some lettuce.

8 The plane left the airport.

The airport was in Spain.

The plane flew to China.

Recognition sentences.

1.1 The dog wanted the dinner.

1.2 The dinner was steak.

1.3 The dog wanted the steak.

1.4 The dinner was biscuits.

2.1 The car drove over the hat.

2.2 The hat was in the road.

2.3 The car drove in the road.

2.4 The hat was in the garage.

3.1 The cow stood next to the sheep.

3.2 The sheep was in the farm.

3.3 The cow was in the farm.

3.4 The sheep ate some grass.

4.1 The bike crashed into the van.

4.2 The van was near the playground.

4.3 The bike was near the playground.

4.4 The van skidded on the gravel.

5.1 The crow was in a field.

5.2 The cat hid in a bush.

5.3 The cat was in a field.

5.4 The crow hid in a bush.

6.1 The video was in the living room.

6.2 The TV smashed to pieces.

6.3 The TV was in the living room.

6.4 The video smashed to pieces.

7.1 The food was carrots.

7.2 The rabbit looked for some lettuce.

7.3 The rabbit ate the carrots.

7.4 The food was some lettuce.

8.1 The airport was in Spain.

8.2 The plane flew to China.

8.3 The plane left Spain.

8.4 The airport was in China.

Story set B

Stories.

1 The doctor sat on the bench.

The bench was in the park.

The doctor felt hot.

2 The book lay on the table.

The table was in the library.

The book came from Italy.

3 The lady ate the lunch.

The lunch was a salad.

The lady wanted a sandwich.

4 The box contained the ball.

The ball was very big.

The box fell apart.

5 The boy sat next to the girl.

The girl was crying.

The boy called a teacher.

6 The police-car chased the truck.

The truck was on the motorway.

The police-car flashed its lights.

7 The man rode the horse.

The horse was running fast.

The man went through the forest.

8 The bus left the town.

The town was in England.

The bus drove to Wales.

Recognition sentences.

1.1 The doctor sat on the bench.

1.2 The bench was in the park.

1.3 The doctor was in the park.

1.4 The bench felt hot.

2.1 The book lay on the table.

2.2 The table was in the library.

2.3 The book was in the library.

2.4 The table came from Italy.

3.1 The lady ate the lunch.

3.2 The lunch was a salad.

3.3 The lady ate the salad.

3.4 The lunch was a sandwich.

4.1 The box contained the ball.

4.2 The ball was very big.

4.3 The box was very big.

4.4 The ball fell apart.

5.1 The girl was crying.

5.2 The boy called a teacher.

5.3 The girl sat next to the boy.

5.4 The boy was crying.

6.1 The truck was on the motorway.

6.2 The police-car flashed its lights.

6.3 The police-car was on the motorway.

6.4 The truck flashed its lights.

7.1 The horse was running fast.

7.2 The man went through the forest.

7.3 The horse went through the forest.

7.4 The man was running fast.

8.1 The town was in England.

8.2 The bus drove to Wales.

8.3 The bus left England

8.4 The town was in Wales.

UNIVERSITY
OF BRISTOL